Introductory Physics

Second Edition

Robert Karplus

UNIVERSITY OF CALIFORNIA
BERKELEY

Introductory Physics

Edited by Fernand Brunschwig
EMPIRE STATE COLLEGE
NEW YORK

a model approach

Second Edition

Introductory Physics: A Model Approach, Second Edition

Editor of Second Edition: Fernand Brunschwig

ISBN 0-9721726-1-0
Printed in the United States of America by Lightning Source, Inc.

Printed copy of Second Edition is available from ABEbooks.com, Amazon.com and other sources on Internet.

To Betty

Foreword

Robert Karplus was my graduate advisor and mentor from 1968 to 1972 at the University of Calfornia, Berkeley. *Introductory Physics: A Model Approach* was just being published when I first met Karplus, and I assisted in teaching several courses based on the book. *Introductory Physics* made a very strong impression: there was Karplus' unique approach to physics, his brilliant insight, his clear, direct language, and his powerful teaching methods, all expressions of his joyful personality, creativity, and love of humanity. However, soon after I finished my Ph. D. and left Berkeley, the book went out of print. Karplus and his wife Elizabeth bought the rights back from the publisher in 1980, and he hoped to publish a second edition. But this was not to be, and he passed away in 1990.

Recently, Robert G. Fuller edited and published a wonderful collection of Karplus' work on science education, along with essays about him by many of his closest collaborators: *The Love of Discovery* (Kluwer Academic/Plenum Press, 2002). Reading the essays, especially the ones by Alan Friedman, Rita Peterson and Fuller, reminded me about how inspiring and influential Karplus was for me and for so many others. More pointedly, re-reading the two chapters of *Introductory Physics* included in the collection made me realize, even more than when I was in Berkeley, how unusual the book was and how valuable it still could be. I resolved to try to bring *Introductory Physics* back to life.

My initial thought was to reprint the 1969 edition of the book. But this would not really have been in Karplus' spirit. He would have wanted to fix errors and clarify the text wherever possible. At first, I hesitated to change anything, but I found that I could indeed distinguish what really needed change from the many valuable features that had to be preserved. This required a substantial effort – more than a year – but it has been a labor of love.

With the permission and encouragement of Elizabeth Karplus, I have edited *Introductory Physics*, and this second edition is published in memory of Robert Karplus.

Robert Karplus was a well-known physicist, physics teacher and science educator: Professor at UC Berkeley, Director of the Science Curriculum Improvement Study (SCIS), Director of Berkeley's Lawrence Hall of Science, President of the American Association of Physics Teachers, and a creative, prolific researcher in physics and science education. In addition, Karplus was a warm, generous person, a wonderful teacher, a terrific speaker, a passionate advocate for good science teaching, and a tireless campaigner for innovation and improvement in science education at all levels.

After more than ten years at the top echelon of theoretical physics research, Karplus became interested and then immersed in science education. His subsequent career spanned a unique period – the Post-Sputnik effort to improve understanding of science.

In this long effort, Karplus was a pioneer: he was one of the very first (possibly the first) leading scientists to focus on elementary science education, and he was instrumental in extending the national priorities, and funding, which focused first on the college level then on the secondary level and finally, after Karplus, on the elementary level. He realized the importance of Piaget's work very early, became a vocal expert on Piaget's ideas, and did significant original research on the reasoning process.

Karplus originated the Science Curriculum Improvement Study in 1963, and over the next 15 years, in collaboration with Herbert Thier and many others, he carried through a major national elementary-level curriculum reform effort. Karplus, Thier and SCIS brought good science (including hands-on investigations) into schools throughout the country and beyond. With a deep-seated belief in the values of public education and democracy, and a self-evident spirit of cooperation and goodwill, SCIS made a very significant contribution to elementary education in the US.

In fact, Karplus was one of the most effective and charismatic leaders in the many-faceted effort of the 1960s and 70s to develop science curricula, train teachers, carry out research, educate voters and political leaders, and generally upgrade the understanding and teaching of science throughout the US. While he is best known for his work on SCIS at the elementary level, Karplus taught physics regularly at Berkeley, he was active in the American Association of Physics Teachers, and he made many significant contributions to physics teaching at the college level. *Introductory Physics* is the fruit of these efforts.

A longer biography of Robert Karplus appears at the end of the book. I would like to express my gratitude to Elizabeth Karplus for her permission to revise and publish the book, and especially for her strong and nurturing encouragement along the way. I would also like to thank Empire State College (where I have taught since 1972), Donald S. Cook of the Bank Street College of Education, Alan Friedman of the New York Hall of Science, my wife Jennifer Herring, my publisher Timothy Johnson (a distinguished graduate of Empire State College), Harold Ohrbach of the New York Academy of Sciences, and the various individuals, publishers and organizations listed in the Acknowledgments for their permission to include illustrations, diagrams or photographs in the Second Edition.

Fernand Brunschwig

fbrunsch@esc.edu
New York, New York
May, 2003

Introduction to the Second Edition

Robert Karplus wrote this innovative textbook on physics for non-science students between 1965 and 1969. The book first appeared in 1966 as a preliminary edition, and in 1969 W. A. Benjamin, Inc. published the first edition in hardback. For reasons not connected with the quality of the book, it never got to a second edition and has been out of print for many years.

As Karplus says in the Author's Preface below, ". . . this book is addressed particularly to readers with little scientific or mathematical background: the only requirements are common sense, experience, and reasoning ability." Writing an understandable physics textbook for readers with little background is a laudable goal – often announced, but seldom achieved. In fact, a large variety of magazines, popular books, and textbooks have been published for this audience since 1969. So, you ask, why republish *Introductory Physics: A Model Approach* now?

The reason is that this is a genuinely outstanding physics textbook - understandable to students and rewarding to teach. Whether you are a physics student, a teacher, a prospective teacher, or a general reader, you can benefit from the many insights and innovations that Karplus poured into *Introductory Physics*. As in all his work, you will find here an original attack on the subject at the most fundamental level expressed in clear, direct language with an extraordinary sense of the joy of discovery.

The feature of *Introductory Physics* that is most valuable to students and teachers, and which sets it off from most other physics texts, is that Karplus really does start at a beginner's level. In fact, the first four chapters (Part One, a full quarter of the book) are a carefully graduated "on ramp," a *tour de force* of physics teaching.

Karplus and his team at SCIS had worked hard and long to create, test and revise an entire curriculum (including a sequence of topics and activities exemplifying Piaget's ideas, plus an imaginative set of hands-on investigations grounded in children's experiences). Karplus took full advantage of all this; Part One of *Introductory Physics* focuses on the topics he knew to be essential: reference frames, relative motion, working and mathematical models, systems and subsystems, interaction, radiation, and energy. In his unique way, Karplus develops these ideas in Part One, drawing on many illustrations and concrete examples from SCIS plus his mastery of Piaget's theory, to build the reader's concrete and abstract reasoning skills.

In Chapters 3 and 4, Karplus' extraordinary capacity to see physics as a whole, and his skills in simplifying and synthesizing, come into play: he focuses on the concepts of interaction, fields, radiation and energy, drawing upon gravity, electricity, magnetism, heat, and many other phenomena as examples to provide an accessible, powerful, integrated and satisfying overview of the field. By the end of Chapter 4, the student has some real experience with the process of understanding the natural world and is familiar with what physics involves, how it is done, the key concepts and terminology, and its rewards and shortcomings.

Beyond its value as a textbook, *Introductory Physics* provides a window through which we can see how a master synthesized his expertise in physics, psychology, and education, and how he built on the innovations generated by the dynamic curriculum and teaching reform projects of the time. Naturally, as described above, he drew upon his 15 years of work with elementary school children and elementary science curricula to create Part One of the book. Another very important influence on Karplus' thinking about how to make physics understandable to a beginner was Francis Friedman, the convenor of the Physical Science Study Committee (PSSC) at MIT in 1956. By 1969, the PSSC, now under the direction of Jerrold Zacharias and with ongoing support from the National Science Foundation, had generated the first and second editions (1960 and 1965) of the innovative *PSSC Physics Course* for secondary schools.

PSSC Physics was based on Friedman's willingness to face the fact that velocity, acceleration, force and Newton's laws of motion, though the starting point for all physics courses of the time, actually demanded a level of abstract thinking for which most students were not prepared. Friedman identified other topics, particularly the study of light (ray and wave optics), as requiring substantially less abstract thinking. Therefore, *PSSC Physics* reversed the conventional order of topics, using the study of light, especially a critical comparison of the wave and particle models, plus other topics, to build up the reasoning skills needed to understand Newtonian mechanics.

Karplus enthusiastically adopted this approach, which fitted well with his own view of physics as well as his research on the development of reasoning. Karplus had also spent more than 10 years on the forefront of theoretical quantum field theory; he was intimately familiar with the difficulties of trying to apply the classical Newtonian approach in modern physics, and he was very aware of the on-going value of concepts such as interaction, radiation, field, momentum, and energy. This conceptual background very likely also made Karplus especially receptive to Friedman's idea.

Karplus also built on the PSSC comparison of the wave and particle models. But, in Karplus' hands the study of scientific models became the main theme. In Part One, he explains many trenchant examples illustrating how models are built, adapted and discarded. In Part Two (Chaps. 5-8), he compares the wave and particle models for light and (in distinction with PSSC) also for sound. Chapter 8 ("Models for Atoms") represents the climax of the first half of the book. Karplus uses the wave model (rather than force and Newton's laws) as his primary tool to elucidate modern physics. He succeeds brilliantly here: allowing the concepts to grow naturally from their context and explaining ideas such as the Bohr model, electron diffraction, wave mechanics, the wave-particle duality and the uncertainty principle in a way that makes sense to beginners.

Karplus' teaching approach was also strongly influenced by Gerald Holton, a former colleague in graduate school at Harvard. Holton, a physicist and a historian of science at Harvard, had a deep understanding of the relationships between science and other human concerns. In collaboration with Fletcher Watson and F. James Rutherford, Holton developed the *Project Physics Course*, which took a substantially different approach to secondary school physics than PSSC. Karplus was very appreciative of *Project Physics*, which identified ways to teach physics as part of the human experience and integrated the development of scientific concepts within a historical and cultural context. Karplus borrowed some specific ideas from *Project Physics*, for example the attempt to isolate a single light ray in Section 7.2, and he drew upon Holton's wonderful textbook *Development of Concepts and Theories in Physical Science* (1952) as well as Thomas Kuhn's seminal work, *The Structure of Scientific Revolutions,* which came out in 1962. Karplus also clearly did substantial reading of his own in the scientific and historical literature.

There are many notes in the margins about the historical context. More important, there are innumerable points in the course of an explanation of the physics content where Karplus calls up the historical context or the thinking of the discoverer in just the way needed to help a beginner gain a deeper understanding or to relate the concepts together in a memorable way.

This edition of Karplus' book is still quite close to its original form. While I edited the text in many ways – to clarify key points, correct typos, and modernize the most glaringly out-of-date items – I kept the page layout as close to its original form as possible. This imposed a rather demanding, yet salutary, requirement: modifications had to fit within the same number of pages as the original. Matching words in this way with the master gave me a workout at a level that I had not experienced since I completed my dissertation under his direction over 30 years ago.

This edition of Karplus' text could well be updated and supplemented in many ways. In particular, the references could be brought more up to date, many valuable innovations in physics teaching of the last 30 years could be incorporated, and additional hands-on experiments could be included. I would hope that the value of Karplus' work will not be dimmed by any shortcomings in the editing or by the fact that it has been out of print for so long.

Is understanding of the fundamental concepts of science necessary for non-scientists today? Is the capacity to apply scientific principles, think logically, and solve problems important for national and world citizenship? To Karplus the answers were obvious. His book provided and, I believe, still provides a positive and realistic response to these on-going challenges. I hope that a new generation of teachers and students find the book as valuable, provocative, and on-target as I, and many others, did during its brief life in print.

Fernand Brunschwig

New York, New York
June, 2003
email: fbrunsch@esc.edu

Author's Preface

What is physics? Every introductory physics text tries to answer this question in its own way, and this text is no different. However, this book is addressed particularly to readers with little scientific or mathematical background: the only requirements are common sense, experience, and reasoning ability. To make the subject meaningful to this audience, it was necessary to alter the customary approach of most introductory texts.

The usual sequence of topics in introductory physics courses for liberal arts students begins with a study of Newtonian mechanics and its roots in astronomy. There is good reason for starting with mechanics, because Isaac Newton's brilliant achievement three hundred years ago marked the beginning of the interplay of experiment, imagination, and deductive logic that has characterized physics ever since. Unfortunately for the novice, Newton's imaginative leap from observation to formulation of axioms was so great that few can follow it. Introducing Newtonian mechanics at the outset of the course presents an extremely difficult problem for many students. This problem often acts as a deterrent to their full appreciation of physics since many students never realize that other subject areas (for example, heat and temperature, sound, light, elasticity) do not make the same demands for abstract reasoning when they are treated at an elementary level.

To remove this impediment, I have rearranged the subject matter and have placed Newtonian physics at the end of the text. The result is an emphasis on the "interaction" and "energy" concepts rather than on the more abstruse "force" concept; thus, physics is related more effectively to the common sense views of most students entering a beginner's course. The instructor accustomed to the more usual approach will have to be careful to remain within the conceptual sequence of this text.

The book is primarily intended to meet the requirements of a one-semester course for non-science students. It consists of four parts, an epilogue, and an appendix reviewing the mathematical background for the course. Part One is an overview of physical phenomena, which are taken up in more detail in the later parts. A number of fundamental concepts are introduced: scientific models, reference frames, systems, interaction, matter, energy, equilibrium, steady state, and feedback. This initial part lays a foundation for the remainder of the text, and it will be helpful to any beginning student of science whether he continues the study of physics or moves into other areas.

Part Two is concerned with light, sound, radiation, and the structure of atoms. This part introduces the student to twentieth century physics and will enable him to appreciate the many excellent books on relativity theory and quantum physics listed in the bibliographies. In Part Three, operational definitions of energy are used to construct simple mathematical models for thermal energy (specific heat), elastic energy, gravitational field energy, and electric power. Finally, Part Four deals with the Newtonian theory of particle motion relating force to acceleration. Immediate applications of the theory lead to a mathematical model for kinetic energy, to a description of the motion of falling objects, and to an understanding of the action of rocket engines. More extensive investigations include the study of periodic motion (for example, the solar system and the pendulum), gas pressure and heat engines, and the kinetic theory of gases. In this final part, the extensive chapter-end bibliographies will be particularly helpful in guiding the student to further study of topics that interest him most.

It is clear from this survey of the text that its overall organization does not follow the historical development of physics. As previously mentioned, I have been motivated instead by a desire to construct a bridge between the preconceptions of the nonscientific reader and the concepts and theories of physics. In order to provide a helpful and interesting historical context for some aspects of the subject, however, the margins include historical background, biographical notes, and quotations from the works of prominent scientists. Much of this material illustrates the transient nature of most scientific theories and the extensive modifications that scientific models have undergone. What is esteemed by one generation may be discarded by the next!

Although the present sequence of topics is the one I consider most fruitful, the organization of the text into four separate parts (the latter three of which represent fairly independent entities) does allow some flexibility in the order in which these topics may be treated. It is possible, for example, to omit Part Two or to exchange the order of Parts Two and Three. Through such a modification the needs of particular courses or the preferences of the instructor may be met.

The text may also serve as a full year's course. Parts One and Two could be studied in the first semester, and Parts Three and Four in the second semester. To supplement the text material and to expand the course to fit the requirements of a full year's program, selections from the extensive bibliographies could be the focus of many assignments and discussions. These bibliographies also enable the instructor to adapt the course to encompass the diverse interests of the students.

As a final thought, I suggest that the reader begin with a reading of the Epilogue and conclude with Chapter 1. I do not mean that the book should be read backwards. After the Epilogue, the book may be read in the proper sequence, but the reader should return to Chapter 1 after he has read the Epilogue for a second time. By comparing his reaction to these two sections of the book on the first and second reading, he will be able to gauge his progress in scientific literacy.

My work on this book has benefited from the assistance and encouragement of many colleagues, friends, and students, and it is a pleasure to acknowledge them. Abraham Fischler, now at Nova University, teased me into teaching an "impossible" physics course for liberal arts

students and thereby indirectly initiated the project. Thomas Kuhn, Gerald Holton, and the late Francis Friedman of the Physical Science Study Committee strongly influenced my view of what is important in physics. My associates in the Science Curriculum Improvement Study, especially Carl Berger, Joseph Davis, Chester A. Lawson, Marshall Montgomery, Luke E. Steiner, Laurence Strong, and Herbert D. Thier, helped me in selecting and refining my approach to the subject matter. Stephen Williams and Michael von Herzen assisted in the preparation and evaluation of a preliminary edition of this text. Diane Bramwell and my wife Elizabeth contributed to the production of the present edition.

I wish to acknowledge also the assistance of many publishers, organizations, and individuals who have made figures and photographs available to me. A detailed listing appears below in the Acknowledgments.

Two large groups of young people have contributed greatly to the preparation of this work. One of these consists of my physics students at Berkeley, who responded sometimes with enthusiasm and sometimes with despair to my teaching and to a preliminary edition. The second of these consists of the many elementary school children (including Beverly, Peggy, Richard, Barbara, Andy, David, and Peter Karplus) in whose classes I have had the privilege to teach. The knowledge gained from the feedback from both these groups has been of immeasurable value.

Finally, the current spirit of innovation in education, which has been fostered by the National Science Foundation during the past decade, has provided the indispensable background for my undertaking.

Robert Karplus

Berkeley, California
January, 1969

Have you ever sorted the books in your library according to their subject matter, only to find a few remaining that "didn't fit"? In a way, this problem is similar to problems that face a scientist. For example, a scientist collects data on crystals or atomic particles or orbiting planets and must face the fact that some of the data does not fit expectations. Such an experience can be unsettling, but it can also lead to new understanding and insight.

One of the primary objectives of this text is to introduce you to a few of the powerful interpretations of natural phenomena used by the physicist to help organize experience. The text discusses some of these phenomena and the patterns of behavior they exhibit. You, in turn, are asked to examine your own experience for additional data to support or contradict these ideas. Occasionally, an unexpected outcome may compel you to reorganize your thinking. A critical approach to all aspects of the text is in order.

Unfortunately, modern culture has become fragmented into specialties. Science was once a branch of philosophy. In modern times, however, science, especially physics, is no longer an intellectual discipline with which every educated person is familiar. There are many reasons for this state of affairs (Fig. 1.1). Probably the most important is that many individuals do not feel a need for a formal study of nature. They develop a commonsense "natural philosophy" as a result of their everyday experiences with hot and cold objects, moving objects, electrical equipment, and so on. For most people, this seems quite adequate.

A second reason is that many of the questions with which modern physicists are concerned seem remote from everyday life. Physicists now study sub-nuclear particles, matter at ultra-low or extremely high temperatures, cosmic-sized objects such as galaxies, the beginning of the universe, and other extraordinary phenomena. The physics that is accessible to the beginning student has a cut-and-dried aspect that lacks the excitement of a quest into the unknown. Therefore, many students tend to think of physics as a finished story that must be memorized and imitated, rather than as a challenge to the creative imagination.

"Why does this magnificent ... science, which saves work and makes life easier, bring us so little happiness? The simple answer runs – because we have not yet learned to make a sensible use of it."

Albert Einstein

A third reason is the frequently indirect nature of the evidence on which physicists base their conclusions. As a result of this indirect evidence, experimental observations are related to theoretical predictions only through long and complicated chains of reasoning, often of a highly mathematical kind.

A fourth reason, of relatively recent origin, is that science has become identified with the invention of destructive weapons (the atomic bomb and biological warfare) and technological advances whose byproducts (smog, detergents) threaten our natural environment. Many individuals reject science, and especially physics, as alien to sensitive, imaginative, and compassionate human beings.

In this text we will try to overcome those difficulties. We will limit the diversity of topics treated, make frequent reference to the phenomena of everyday experience, and examine carefully the ways in which

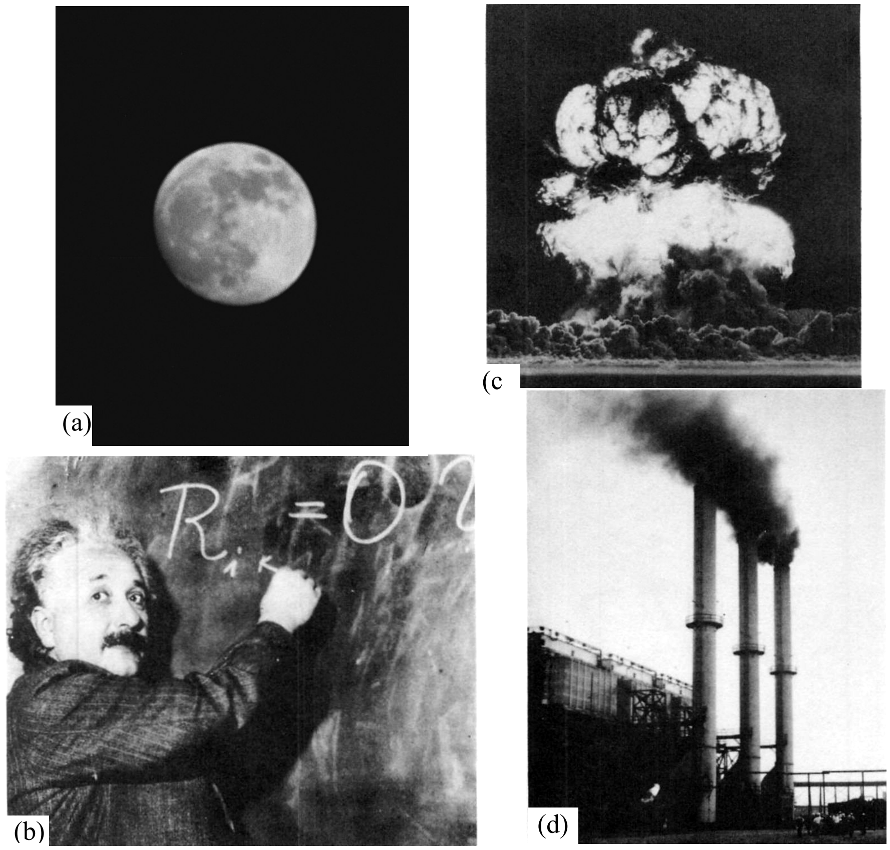


Figure 1.1 What do you think of these reasons?
(a) The familiar surface of the moon. Should you know more about this silvery disk where people may someday live?
(b) Albert Einstein, popular symbol of theoretical, abstruse physics. Should grants for pure research be justifiable in terms of contemporary social needs?
(c) Nuclear explosion, Nevada Proving Grounds, 1957. Physics has become deeply involved in war and peace.
(d) A steel mill's waste gases bring with them one of the hidden costs of our technological civilization. Compare with Fig. 1.2(c).

observations can be interpreted as evidence to support various scientific theories. The goal is to develop your understanding of how physical concepts are interrelated, how they can be used to analyze experience, and that they are employed only as long as there are no better, more powerful alternatives.

The reasons why an educated person should have some understanding of physics have been stated many times (Fig. 1.2). Physics is a part of our culture and has had an enormous impact on technological developments. Many issues of public concern, such as air and water pollution, industrial energy sources, disarmament, nuclear power plants, and space exploration, involve physical principles and require an acquaintance with the nature of scientific evidence. Only a wider public understanding of science will ensure that its potential is developed for our benefit rather than devoted to the destruction of civilization. More personally, your life as an individual can be enriched by greater familiarity with your natural environment and by your ability to recognize the operation of general principles of physics everyday, such as in children swinging and hot coffee getting cold.

1.1 The scientific process

The present formulation of science consists of concepts and relationships that humankind has abstracted from the observation of natural phenomena over the centuries. Throughout this overall evolutionary process occasional major and minor "scientific revolutions" (or, *"Few things are more be-nighting than the condescen-sion of one age for another."*

Woodrow Wilson

possibly more accurately, "transformations") have reoriented entire fields of endeavor. Examples are the Copernican revolution in astronomy, the Newtonian revolution in the study of moving objects, and the introduction of quantum theory into atomic physics by Bohr. The net result has been the development of the conceptual structure and point of view with which modern scientists approach their work.

What happens to an object released in space, far from the earth or another body?

An investigation. Let us briefly and in an oversimplified way look at the way a scientist might proceed with an investigation. For instance, consider a ball that falls to the ground when you release it. After additional similar observations (other objects, such as pieces of wood, a feather, and a glass bowl, all fall to the ground when released), we are ready to formulate a hypothesis: all objects fall to the ground when released. We continue to experiment. Eventually, we release a helium-filled balloon and find that instead of falling, it rises. That is the end of the original hypothesis. Can we modify it successfully? We could say, "All objects fall to the ground when released in a vacuum." This statement is more widely applicable, but it is still limited to regions near the earth or another large heavenly body where there is a "ground." In space, far from the earth, "falling to the ground" is meaningless because there is no ground.

This simple description has skipped over two important decisions that we made. First was the judgment as to what constituted "similar"

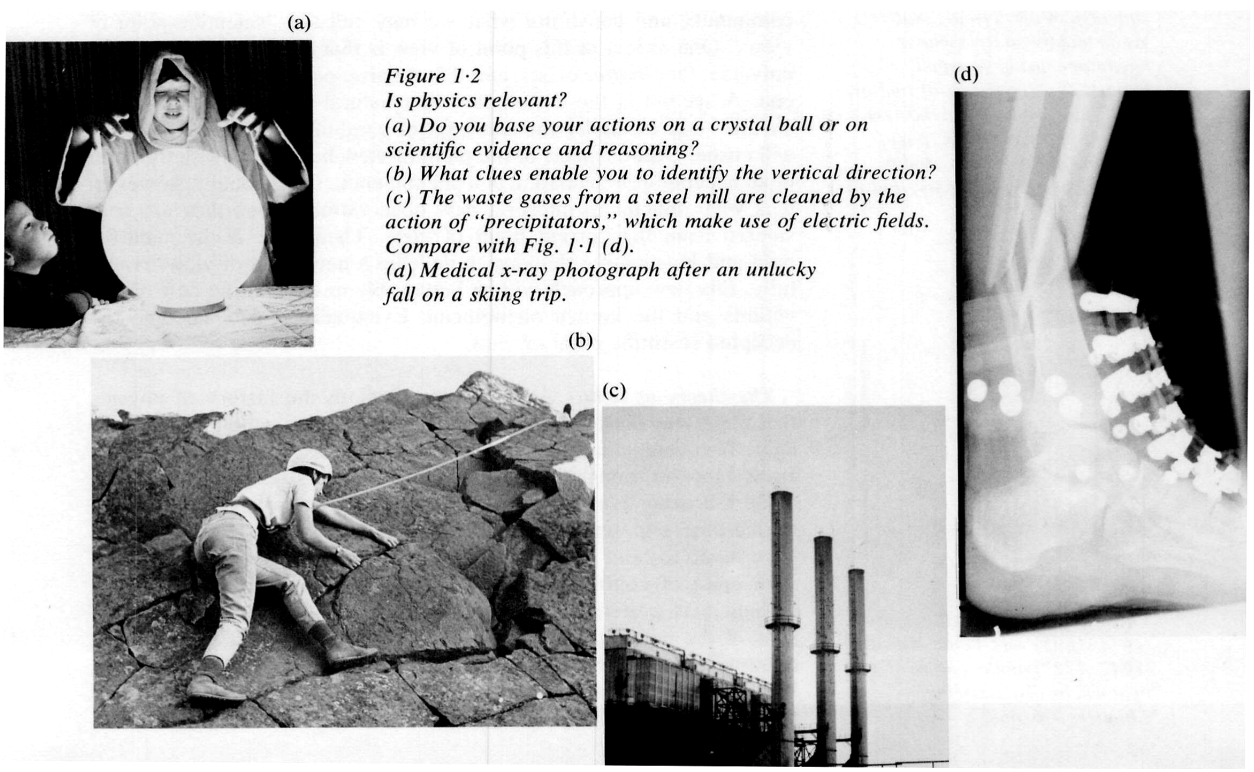
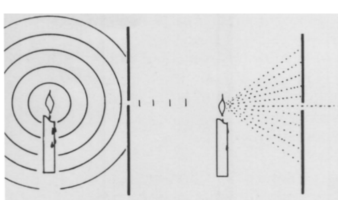


Figure 1·2
Is physics relevant?
(a) Do you base your actions on a crystal ball or on scientific evidence and reasoning?
(b) What clues enable you to identify the vertical direction?
(c) The waste gases from a steel mill are cleaned by the action of "precipitators," which make use of electric fields. Compare with Fig. 1·1 (d).
(d) Medical x-ray photograph after an unlucky fall on a skiing trip.

". . . from my observations, . . . often repeated, I have been led to that opinion which I have expressed, namely, that I feel sure that the surface of the Moon is not perfectly smooth, free from inequalities and exactly spherical, as a large school of philosophers consider with regard to the Moon and the other heavenly bodies, but that, on the contrary, it is full of inequalities, uneven, full of hollows and protuberances, just like the surface of the Earth itself, which is varied everywhere by lofty mountains and deep valleys."
Galileo Galilei
Sidereus Nuncius, 1610

"Matter" includes all solid, liquid, and gaseous materials in the universe. In this text, we will not define "matter" more precisely; we will treat "matter" as an undefined term, with a meaning that must be grasped intuitively. Properties of matter, to be described later in this text, include mass, extent in space, permanence over time, ability to store energy, elasticity, and so on.



The work of Christian Huygens (1624-1695) and Isaac Newton (1642-1727) on the nature of light will he discussed in Chapters 5, 6, and 7.

observations. For instance, in the example, we included the balloon along with the ball, weed, feather, and so on. Yet we might have considered the balloon to be very different from the other objects observed. Then the balloon rising rather than falling would not have been considered pertinent to the hypothesis of falling objects. Even for some time after Galileo's telescopic observations of the moon more than 300 years ago, there was controversy as to whether it and other heavenly bodies were material objects to which the hypothesis of falling objects should apply.

The second decision was the judgment about what aspects of the observations were to be compared. We decided to compare the motion of the bodies after they were released. Aristotle, a Greek philosopher who also thought about falling bodies, was more concerned with such objects' ultimate state of rest on the ground, and therefore he reached conclusions very different from those we found above.

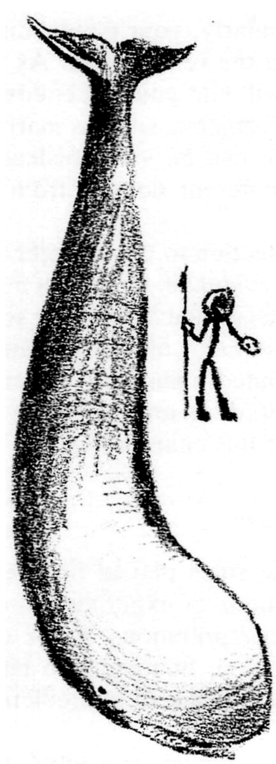
The scientific point of view. Usually the answers to these two kinds of questions are tacitly agreed upon by the members of the scientific community and constitute what we may call the "scientific point of view." One aspect of this point of view is that a real physical universe composed of matter exists, that we are a part of this universe, and that matter participates in natural phenomena. A second is the assumption that natural phenomena are reproducible: that is, under the same set of conditions the same behavior will occur. A third aspect is that while we ourselves are part of the physical world, we are also able to observe the natural world and to think about our observations. Other aspects of the point of view have to do with the form of an acceptable explanation of a phenomenon. This scientific point of view provides a context for scientific knowledge and for what is (and is not) accepted as scientific knowledge. Occasionally, however, it is very difficult to interpret new observations in a way that is consistent with the accepted scientific way of thinking. Then there is the need for bold and imaginative thinking to develop a new point of view. Hopefully, this new approach will be better able to explain the new observations and the known phenomena. Eventually it may become the accepted scientific point of view. The key idea here is that the scientific point of view (that is, the criteria for what is scientific knowledge) has gradually changed and is certain to continue to change.

The theory of light. A fascinating story in the history of physics that illustrates these remarks deals with the nature and interactions of light. Two competing ideas were advanced in the seventeenth century. Isaac Newton thought that light consisted of a stream of corpuscles, while Christian Huygens believed that light was a wave motion (see illustration to left). Up to that time, experiments and observations on light rays had apparently been made without questioning further the nature of the rays.

In spite of contradictory evidence, Newton's corpuscular theory of light was preferred by the scientific community, largely because of the success of Newton's laws of the motion of material bodies subject to forces. Small bodies (corpuscles) probably provided a more acceptable explanation to Newton's contemporaries and followers than did the waves proposed by Huygens. During the nineteenth century, however, new experimental data on the passage of light near obstacles and through transparent materials contradicted Newton's corpuscular theory conclusively and supported the wave theory. Waves and their motion became the accepted way to explain the observed properties of light.

This point of view flourished until the beginning of the twentieth century, when results of further experiments on the absorption and emission of light by matter conflicted with the wave theory and led to the presently accepted quantum theory of light. Already, however, there are contradictions within this theory, so that it, too, will have to be modified. This is one field of currently active research, and several proposals for new theories are being studied intensively to determine which holds the most promise.

Scientific "truth." Science is, therefore, never complete; there are always some unanswered questions, some unexpected phenomena. These may eventually be resolved within the accepted structure of science, or they may force a revision of the fundamental viewpoint from which the phenomena were interpreted. Progress in science comes from two sources: the discovery of new phenomena and the invention of novel interpretations that illuminate both the new and the well-known phenomena in a new way. Scientific truth is therefore not absolute and permanent: rather, it means agreement with the facts as currently known. Without this qualification, the statement that scientists seek the truth is misleading. It is better to say that scientists seek understanding.



1.2 Domains of magnitude

When and how does a person's experience of space and time originate? Probably the foundations are laid before birth, but the most rapid and important development takes place during an infant's early exploration of the environment. By crawling around, touching objects, looking at objects, throwing objects, hiding behind objects, and so on, an infant forms simple notions of space. By getting hungry and feeling lonely, by enjoying entertainment and playing, by watching things move and by moving himself, he forms notions of time. Even though an adult commands more effective skills with which to estimate, discriminate, and record space-time relations, our need to relate the environment to ourselves is never really outgrown.

Size. As you look about and observe nature, you first recognize objects, such as other people, trees, insects, furniture, and houses that are very roughly your own size. We will call the domain of magnitude of these objects the *macro domain*. It is very broadly defined and spans living creatures from tiny mites to giant whales. All objects to which you relate easily are in this domain.

All other natural phenomena can be divided into two additional domains, depending on whether their scale is much larger or much smaller than the macro domain. The former includes astronomical objects and happenings, such as the planet earth, the solar system, and galaxies. We will call this the *cosmic domain.* Much smaller in scale than the macro domain is the one that includes bacteria, molecules, atoms, and subatomic units of matter; we will call it the *micro domain.*

*The phrase "*geologic times*" is sometimes used to denote very long time intervals because geologic processes (such as changes in the shape of the Earth) are extremely slow.*

Time. It is useful to introduce the concept of domains into time scales as well as into physical size. Thus times from seconds or minutes up to years are *macro times* in the sense that they correspond to the life spans of human beings and other organisms. Beyond centuries and millennia are *cosmic times,* whereas *micro times* are very small fractions of a second. As with physical sizes, the mental images you make for processes of change always represent in seconds or minutes what really may require cosmic times or only micro times to occur.

Applications. In order of size, then, the three domains are the micro, macro, and cosmic. The division is a very broad one, in that the earth and a galaxy, both in the cosmic domain, are themselves vastly different in scale. Likewise, bacteria and atomic nuclei are vastly different. Nevertheless, the division is useful because the mental images you make of physical systems are always in the macro domain, where your sense experience was acquired. You therefore have to remember that your mental image of a cosmic system, such as the solar system, is very much smaller than the real system. Similarly, your mental image of a micro system is very much larger than the real system. As you make mental images of these systems, you will find yourself endowing them with physical properties of macro-sized objects, such as marbles, ball bearings, and rubber balls. This device can be very misleading because, of course, your images are in a different domain from the objects themselves.

When we pointed out in the introductory section to this chapter that physicists frequently must interpret indirect evidence, we had in mind, among other things, the three domains of magnitude. Since our sense organs limit us to observations in the macro domain, all interpretations concerning the other domains require extended chains of reasoning. An illustration relating the domains of magnitude to units of space and time measurement is presented at the end of this chapter in Fig. 1.11.

1.3 Theories and models in science

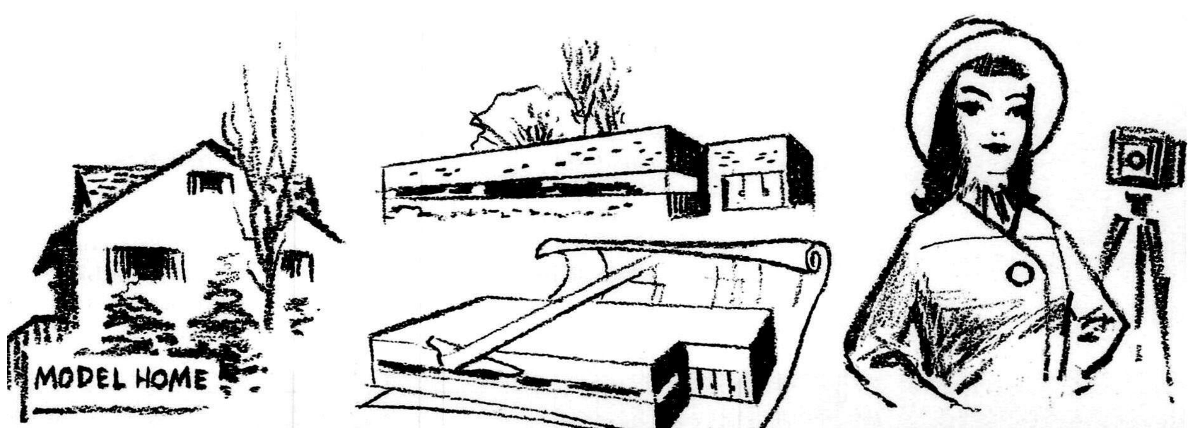
In the preceding section we contrasted the roles played in science by observation and interpretation. Observations of experimental outcomes provide the raw data of science. Interpretations of the data relate them to one another in a logical fashion, fit them into larger patterns, raise new questions for investigation, and lead to predictions that can be tested.

Scientific theories are systematically organized interpretations. Examples are Dalton's atomic theory of chemical reactions, Newton's theory of universal gravitation, Einstein's theory of relativity, and Piaget's theory of intellectual development. Within the framework of a scientific theory, observations can be interpreted in much more far-reaching ways than are possible without a theory. In Newton's theory of gravitation, for instance, data on the orbital motion of the moon lead to a numerical value for the total mass of the earth! In Dalton's theory, the volumes of chemically reacting gases lead to the chemical formulas for the compounds produced. All theories interrelate and extend the significance of the facts that fall within their compass.

Working models. Theories frequently make use of simplified mental images for physical systems. These images are called *working models* for the system. One example is the sphere model for the earth, in which the planet is represented as a uniform spherical body and its topographic and structural complexities are neglected. Another example is the particle model for the sun and planets in the solar system; in this model each of these bodies is represented as a simple massive point in space, and its size as well as its structure is ignored. Still another example is the "rigid body model" for any solid object (a table, a chair) that has a definite shape but may bend or break under a great stress.

Unlike other kinds of models (Fig. 1.3), a working model is an abstraction from reality. Our thoughts can never comprehend the full complexity of all the details of an actual system. Working models are always simplified or idealized representations, as we have already pointed out. Working models, therefore, and the theories of which they are a part, have limitations that must be remembered when their theoretical predictions fail to agree with observations.

Figure 1.3 The word "model" has many connotations in the English language, and most of them are not applicable to the scientific meaning of the word. A scientific "working model" has very little in common with a scale model (model airplane, left), a sample for examination (model home, below left), a visual replica (architectural model, below center), or a person (artist's or fashion model, below right).



The scientist's relationship to the models he constructs is ambivalent. On the one hand, the invention of a model engages his creative talent and his desire to represent the operation of the system he has studied. On the other hand, once the model is made, he seeks to uncover its limitations and weaknesses, because it is from the model's failures that he gains new understanding and the stimulus to construct more effective models. Both creative and critical faculties are involved in the scientist's work with models.

One feature of working models is frequently disturbing to nonscientists: no model perfectly matches reality, and you never know whether a particular model is "right." In fact, the concepts "right" and "wrong" do not really apply to models. Instead, a model may be more or less adequate, depending on how well it represents the functioning of the system it is supposed to represent. Even an inadequate model is better than none at all, and even a very adequate model is often replaced by a still more adequate one. The investigator has to determine whether a particular model is good enough for his purposes or whether it is necessary to seek a better one.

Analogue models. Before a scientist constructs a theory, he often realizes that the system he is studying operates in a way similar to another system with which he is more familiar, or on which he can conduct experiments more easily. This other system is called an analogue model for the first system. You may, for instance, liken the spreading out of sound from a violin to the spreading out of ripples from a piece of wood bobbing on a water surface.

The analogue model for one physical System A is another, more familiar, System B, whose parts and functions can be put into a simple correspondence with the parts and functions of System A. For example, an analogy may be drawn between the human circulatory system and a residential hot water heating system (Table 1.1, below). It is clear that

TABLE 1.1 ANALOGUE MODEL FOR THE HUMAN CIRCULATORY SYSTEM

System A: Human circulatory system	**System B:** Residential hot water heating system
veins, arteries	pipes
blood	water
oxygen	thermal energy
heart	pump
lungs	furnace
capillaries	radiators
hormones	thermostat
(model fails)	overflow tank
(or dilation of veins & arteries)	
blood pressure	water pressure
white blood cells	(model fails)
carbon dioxide	(model fails)
kidneys	(model fails)
intestine	(model fails)

the human circulatory system fulfills several functions, whereas the heating system fulfills only one. The analogue model is, therefore, not complete, but it is nevertheless instructive.

The virtue of an analogue model is that System B is more familiar than System A. This familiarity can have several advantages:

1. Features of the analogue model can call attention to overlooked features of the original system. (Had you overlooked the role of hormones in the circulatory system, the room thermostat would have reminded you.)
2. Relationships in the analogue model suggest similar relationships in the original system. (Furnace capacity must be adequate to heat the house on a cold day; lung capacity must be adequate to supply oxygen needs during heavy exercise.)
3. Predictions about the original system can be made from known properties of the more familiar analogue model. (Water pressure is high at the inflow to the radiators, low at the outflow; therefore, blood pressure is high in the arteries, low in the veins.)

The limitations of the analogue model can lead to erroneous conclusions, however. On a cold day, for instance, the water temperature is higher in the radiators; therefore, you might predict that the oxygen concentration in the blood will be higher during heavy exercise. Actually, the heartbeat and the rate of blood flow increase to supply more oxygen - the oxygen concentration does not change greatly.

"There are two methods in which we acquire knowledge - argument and experiment." Roger Bacon (1214-1294)

Thought experiments. In a thought experiment, a model is operated mentally, and the consequences of its operation are deduced from the properties of the model. A thought experiment differs from a laboratory experiment in that the latter serves to provide new information about what really happens in nature, whereas the former seeks new deductions from previous knowledge or from assumptions. By comparing the deductions with observations in real experiments, you can find evidence to support or contradict the properties or assumptions of the model.

A simple example of a mystery system (Fig. 1.4) can be used to illustrate these ideas. Two working models for what might be under the cover in Fig. 1.4 (a) are shown in Figs. 1.4 (b) and (c). If you conduct simple thought experiments with these models, you quickly find out how satisfactory they are. In the first thought experiment, you imagine turning handle A clockwise. In model G, handle B will turn somewhat faster, because the second gear is smaller than the first, but it will turn counterclockwise. This prediction is in disagreement with the properties of the mystery system. In the second thought experiment, you turn handle A in model S. What can you infer from this second experiment? Can you suggest a satisfactory working model?

Thought experiments are important tools of the theoretical scientist because they enable him to make deductions from a working model or a theory. These deductions can then be compared with observation. The usefulness of a theory or model is determined by the agreement between the deduction and observation. Some very general theories,

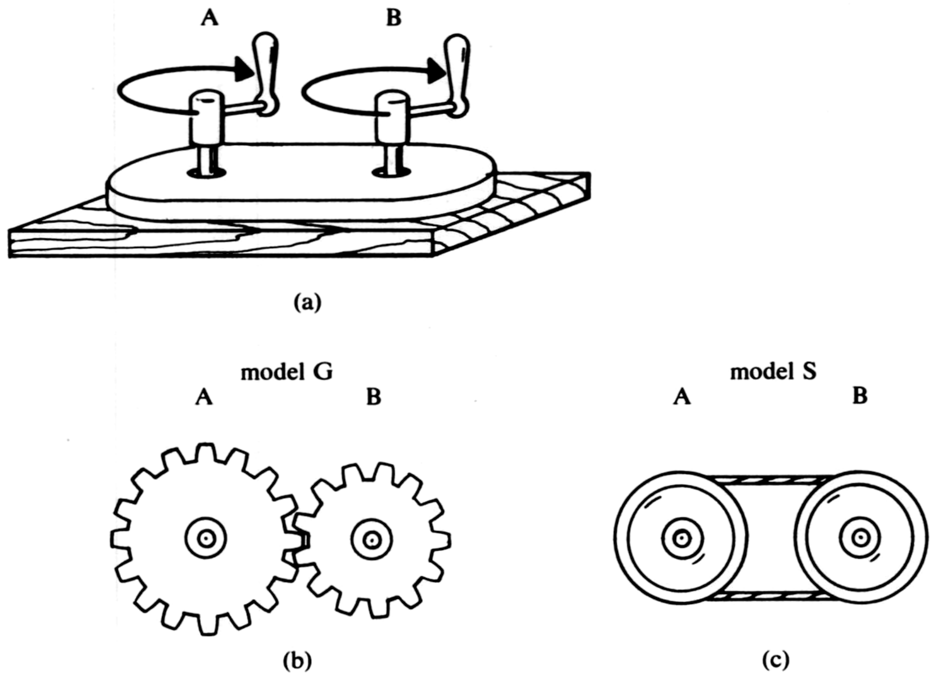


Figure 1·4 A mystery system. (a) When handle A is turned one revolution clockwise, handle B makes 2½ revolutions clockwise. Make models for what is under the cover. (b) Large and small gear model. (c) Two pulley and string model.

Equation 1.1

Mathematical model (algebraic form):
number of turns of handle A = N_A
number of turns of handle B = N_B

$$N_A = N_B$$

Equation 1.2

Mathematical model (algebraic form):
distance $= s$
speed $= v$
time $= t$

$$s = vt$$

such as the theory of relativity, lead to consequences that appear to apply universally. Some models, such as the corpuscular model for light, are useful only in a very limited domain of phenomena.

Mathematical models and variable factors. Scientific theories are especially valuable if they lead to successful quantitative predictions. Working models G and S for the mystery system in Fig. 1.4 both lead to quantitative predictions for the relationship between the number of turns of handles A and B. The relationship deduced from model S (that the handles turn equally) can be represented by the formula in Equation 1·1. We will call such relationships *mathematical models;* the formula in Equation 1.1 is an algebraic way of describing the relationship, which we have also described in words, and which can be described by means of a graph (Fig. 1.5).

A familiar example of a mathematical model, applicable to an automobile trip, is the relation of the distance traveled, time on the road, and speed of the car (Equation 1.2). The distance is equal to the speed times the time. At 50 miles per hour, for example, the car covers 125 miles in 2 ½ hours (Fig. 1.6).

The physical quantities related by a mathematical model are called *variable factors* or *variables*. The numbers of turns of handles A and B are two variable factors in Equation 1.1 and Fig. 1.5. The distance and elapsed time are two variable factors in Equation 1.2 and Fig. 1.6, The speed in this

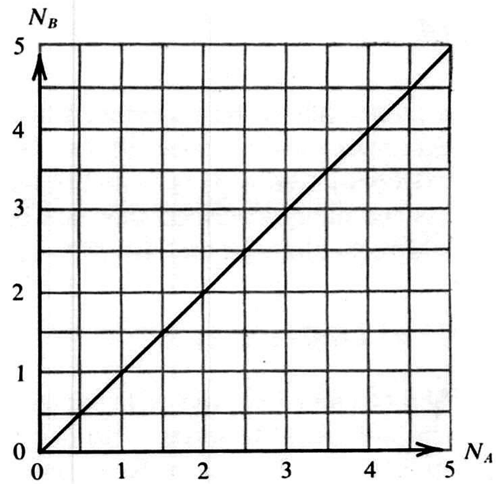


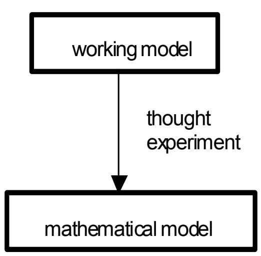
Figure 1.5 Mathematical model (graphical form).

Number of turns of handle A = N_A;

Number of turns of handle B = N_B;

mathematical model is called a constant, because it does not vary. Under different conditions, as in heavy traffic, the speed might be a variable factor.

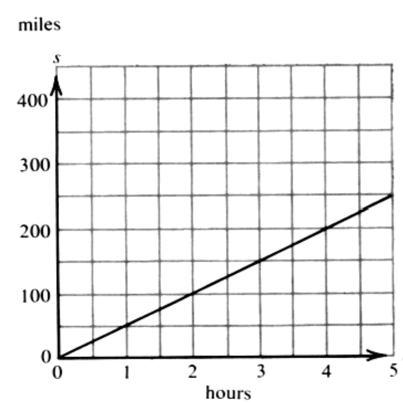
Like the working model for a system, the mathematical model for a relationship is not an exact reproduction of a real happening. No real car, for instance, should be expected to travel at the perfectly steady speed of 50 miles per hour for 2½ hours. The actual speed would fluctuate above and below the 50-mile figure. The actual distances covered at various elapsed times, therefore, might be a little more or a little less than those predicted by the model in Eq. 1.2 and Fig. 1.6. Nevertheless, the model gives a very good idea of the car's progress on its trip, and it is very simple to apply. For these reasons, the model is extremely useful, but you must remember its limitations.



Scientific theories. The making of a physical theory often includes the selection of a working model, the carrying out of thought experiments, and the construction of a mathematical model. All physical theories have limitations imposed by the inadequacies of the working model and the conditions of the thought experiments. Occasionally a theory has to be

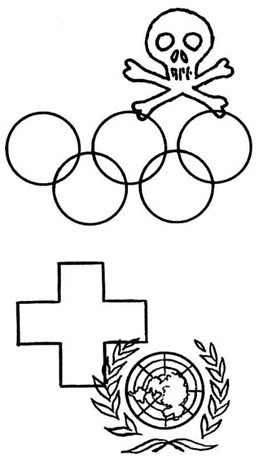
Figure 1.6 Mathematical model of relationship between distance and time (graphical form):

Distance = s (miles),
time = t (hours),
speed = 50 miles per hour.



abandoned because it ceases to be in satisfactory agreement with observations. Nevertheless, physical theories are extremely useful. It is probably the power of the theory-building process we have described that lies behind the rapid progress of science and technology in the last 150 years.

1.4 Definitions



The primary function of language is to communicate information from one individual to others. Human language consists of signs, gestures, spoken sounds, and marks on paper that function as symbols of some sensed experience. Communication by means of human language is possible so long as the communicants have a common understanding of the meaning of the symbols, that is, so long as all persons relate a given symbol to a particular common experience and to none other.

In learning a language, you must first learn to recognize the symbol, then to relate that symbol to a particular experience. A symbol may refer to a material object, the relation or state of material objects, other symbols, or relations of symbols. The normal device for conveying the meaning of a symbol is the definition, of which we will distinguish two types. These are formal definitions, which use words, and operational definitions, which use operations.

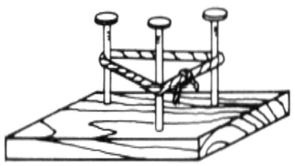
Formal definitions. The familiar dictionary definitions, which identify the meaning of a symbol by the use of words or other symbols, are included in the category of formal definitions. Synonyms, paraphrases, lists of properties, and names of examples are the usual techniques of formal definition.

bottle = jar ?

An example of the use of synonym is "bottle = jar." Synonyms may have exactly the same meaning, but they usually have slightly different meanings. For example, both "bottle" and "jar" are "containers made of glass" (paraphrase) but usually connote different shapes.

An example of definition by paraphrase is "photosynthesis = the conversion of light energy into chemical energy in green plants." Another example is "velocity of an object = the distance traveled divided by the time taken." The paraphrase definition is similar to the definition by synonym, except that the paraphrase contains more words. The paraphrase definition leads to efficiency in communication (or thought, which is self-communication) in that you can substitute the shorter term for the longer phrase. We will occasionally use paraphrases based on a mathematical process to define physical terms, as we did in the velocity example just given.

Operational definitions. The use of real objects and operations (not merely words) to produce, measure, or recognize an instance of a term is the essence of the operational definition. For example, the operational definition of color words, such as red, yellow, mauve, and lime, may be based on a set of color chips that have sample colors on one side and their names on the back. The objects used in this definition are the color chips. The operation is that of comparison of the hue of an unidentified color with those of the color chips. This operational definition is in general use in paint stores.

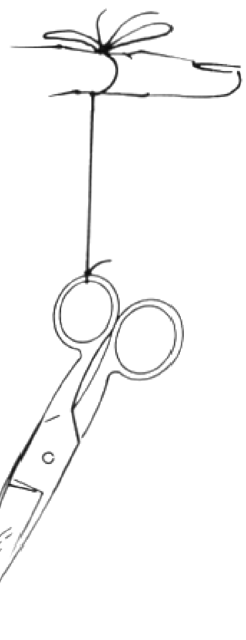


Words can be used to describe an operational definition but the definition itself consists of operations on real objects and not of words. For example, you can construct a triangle by driving three nails into a board and connecting the nails with a stretched string. A figure that matches the figure constructed in this way is also a triangle. These objects and operations define a triangle.

An example of an operational definition that leads to measurement is as follows: the number of seats in an auditorium is the auditorium's capacity. Here the actual seats in the auditorium and the counting operation are combined in an operational definition of the auditorium capacity.

We will soon introduce operational definitions for measuring basic physical quantities, such as length, time, and mass. Each of these definitions makes reference to a standard object that serves as the unit of measurement (in the definition of auditorium capacity, the chair served as the unit of measurement) and a comparison operation that allows the unit to be compared with other objects.

For science, the significance of operational definitions is that their use keeps the description of models and the statement of theories meaningful and testable in the physical world. In contrast to the scientist's operationally defined language, that of the poet rests mainly on terms (for example "beauty," "love," and "grace") that are not defined operationally. However, it is also worth pointing out that the language of a poem generally *does* have a close relationship (or multiple relationships) with the significance, sound, and/or meaning of the words as they are used in the language at large. We also must realize that while scientific concepts must always be somehow logically tied to operational definitions, many scientists use concepts that are only tied to an operational definition through a series of formal definitions. Therefore, scientists often use language that appears just as distant from the real world as the poet's! Finally, poets have anticipated key scientific developments, for example, in ancient times, Lucretius speculated about atoms in his poem, *On the Nature of Things*.



Comparison of formal and operational definitions. In science, formal definitions are frequently used to define one concept in terms of other concepts. For instance, the term "triangle" could have been defined by paraphrase as "a plane figure bounded by three nonparallel straight lines." This definition uses concepts, such as "plane," "nonparallel," "three," and "straight line," for which definitions have to be provided or that may properly remain undefined.

Let us consider another term, "vertical," that can be defined operationally or formally. In the operational definition, a freely hanging plumb line is allowed to come to rest; vertical is the direction indicated by the plumb line. The formal definition is "vertical = the direction toward the center of the earth." The latter definition is a paraphrase that is useful for theoretical purposes, but impossible to apply in practice, as when a house's walls are to be built.

The difference between formal and operational definitions is illustrated especially clearly by their application to "intelligence" and "IQ."

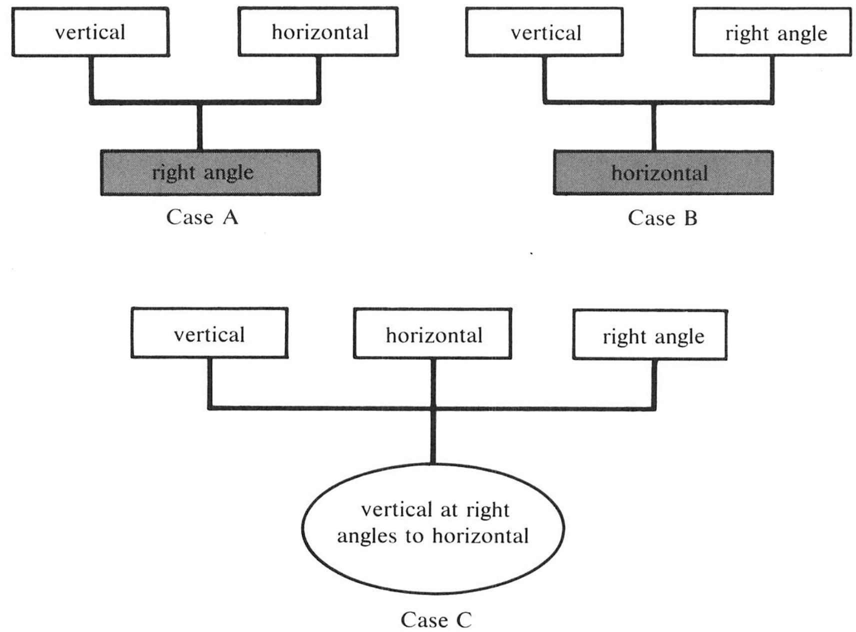


Figure 1.7 Definitions of vertical, horizontal, and right angle. Open box: operationally defined; shaded box: formally defined; oval: experimentally discovered. The definitions are described in Table 1.2

The dictionary defines intelligence as "the ability to apprehend the interrelationships of presented facts in such a way as to guide action toward a desired goal." The value of this formal definition as a positive personal trait seems obvious. It is very difficult, however, to rank individuals according to their intelligence, because this requires applying the definition operationally to specific cases. The intelligence

TABLE 1.2 THREE ALTERNATIVE DEFINITIONS OF VERTICAL, HORIZONTAL AND RIGHT ANGLE

Case A. Define vertical: direction of a free plumb line at rest.
Define horizontal: direction of a free water surface at rest.
Define right angle: the angle between the vertical and horizontal.

Case B. Define vertical: direction of a free plumb line at rest.
Define equal angles: angles that match when superposed.
Define straight line: matches a stretched string.
Define right angle: draw two intersecting straight lines on a given (flat) board so that four equal angles are produced. Each angle is a right angle.
Define horizontal: the surface at right angles to the vertical.

Case C. Define vertical: direction of a free plumb line at rest.
Define horizontal: direction of a free water surface at rest.
Define right angle: draw two intersecting straight lines on a given (flat) board so that four equal angles are produced. Each angle is a right angle.
Experimental relation: vertical and horizontal make a right angle.

quotient (IQ) can be defined operationally by a standard score on a specific test combined with a person's age. However, the *meaning* of the IQ as a personality trait and its functional value (that is, the relationship between an operationally defined IQ and its more generally accepted formal definition) are subjects of controversy that are far from being resolved.

Formal definitions and operational definitions each have their advantages and disadvantages. Operational definitions, as we have already stressed, make direct reference to the physical world and to human perception. This property gives them the advantage of being concrete. At the same time, their dependence on specific objects (such as auditorium seats) limits their scope of application. The definition of "capacity" given for an auditorium, for instance, could not be applied to the gasoline tank of a car. A definition of temperature using an ordinary thermometer would not be applicable in the interior of the sun. Operational definitions tend to be cumbersome in that they demand the availability of certain equipment.

Formal definitions, by contrast, are more concise and efficient. They relate concepts to one another directly. The definitions are much more generally valid. The price that is paid for these advantages is that the language becomes very abstract, because direct connections with reality are buried in the foundations on which the system of formal definitions rests.

In this text we will place more reliance on operational definitions than is customary, because we believe that concrete ties to reality are more valuable to you than efficiency and generality. Our approach, therefore, will be somewhat different from that of other texts. However, the physical world that is being described is the same; the differences are in the logical development and not in the content itself. To illustrate the diversity of possible approaches to the logical development of ideas, Fig. 1.7 and Table 1.2 show how the concepts "vertical," "horizontal," and "right angle" may be defined and related to one another in three different ways.

1.5 Length, time, and mass

That we relate most easily to the macro domain of magnitudes is reflected in the fact that units for measuring length have, since ancient times, been derived from our bodies (Fig. 1.8). The ready availability of the human body made the foot and the inch convenient units, but there was a great deal of local variation, depending on whose foot or thumb was used. With the growth of an international scientific community, it became necessary to adopt standard units of measurement that would be accepted by scientists everywhere. The French Academy of Sciences in 1791 suggested a new unit of length, the meter, which was to be one ten-millionth of the distance from the pole to the equator of the earth. Accordingly, a platinum-iridium bar with two marks separated by the "standard meter" was prepared after seven years of surveying the earth in Spain and France. The original is kept in the Bureau of Weights and Measures near Paris and accurate copies are kept by the National Bureau of Standards near Washington (Fig. 1.9) and by similar agencies

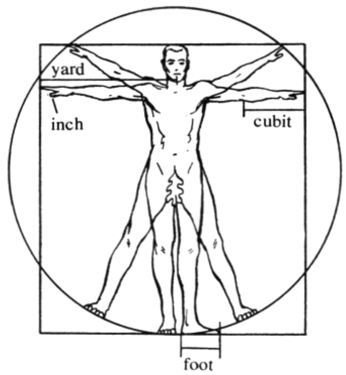


Figure 1-8 Units of measurement related to the human body.

OPERATIONAL DEFINITION

Length or distance is measured by the number of standard units of length that can be placed end to end to match the desired length or distance.

The symbol s (for space) will be used to represent distance.

Abbreviations for units:
1 m = 1 meter
1 cm = 1 centimeter $= 0.01\ m = 10^{-2}\ m$
1 mm = 1 millimeter $= 0.001\ m = 10^{-3}\ m$
1 km = 1 kilometer $= 1000\ m = 10^{3}\ m$

elsewhere. The marks on the rulers you use are derived, through a long chain of copying, from the original standard meter in France.

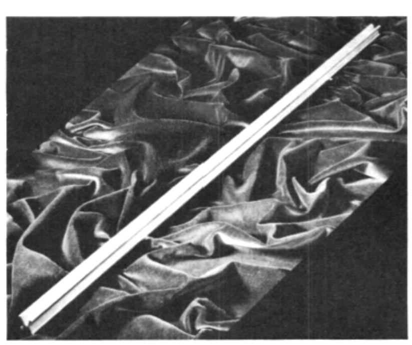
Widely accepted units of measurement are essential to our technological culture. The story of weights and measures and the continuing search for improved standard units will never end.

We turn now to the operational definitions of the basic quantities of length, time, and mass. Since the most primitive measurement operation is that of counting, the definitions involve procedures for comparing the quantity to be measured with accepted standard units and counting the number of standard units that are required.

Length and distance. Length and distance are defined by a matching procedure in which the length of any object can be used as the unit. The generally accepted standard unit of length is the meter, described above. After the meter had been established, it was found that the earlier measurements of the earth had been inaccurate, so the geographical definition was abandoned, but the platinum-iridium bar was kept. However, duplicating the standard length was cumbersome and tended to introduce additional errors. As a result, the current definition of the meter in terms of the wavelength (see Chapter 7) of a specially designed light source was adopted. This definition allows the standard meter to

Figure 1.9 Replicas of the international standards of length and mass. (a) The standard meter bar. (b) The standard kilogram cylinder, whose size is close to that of a small egg.

(a)



(b)

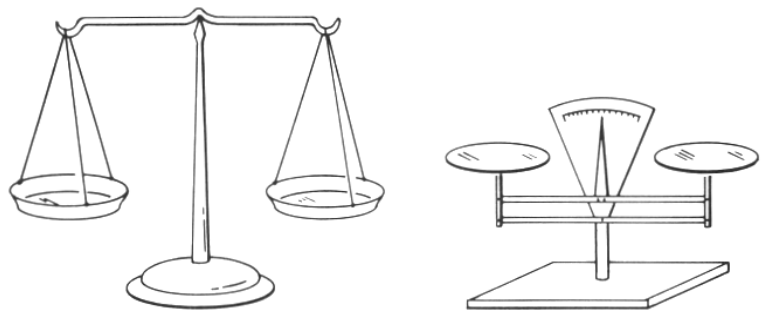


Figure 1.10 Equal-arm balances.

OPERATIONAL DEFINITION
Time interval is measured by the number of standard units of time that elapse during the desired time interval.

The symbol t will be used for time. The symbol Δt will be used for time interval. (Δ is the Greek letter delta, for difference.)
Abbreviations for units:
1 sec = 1 second
1 min = 1 minute = 60 sec
1 hr = 1 hour = 3600 sec
1 day = 86,400 sec
1 yr = 1 year = 3.16 x 10^7 sec

OPERATIONAL DEFINITION
Mass of an object is measured by the number of standard units of mass that are required to balance the desired object on an equal-arm balance.

The symbol M_G will be used for gravitational mass. Abbreviations for units:
1 kg = 1 kilogram
1 g = 1 gram = 0.001 kg = 10^{-3} kg
1 mg = 1 milligram= 10^{-6} kg
1 metric ton= 1000 kg = 10^3 kg
1 megaton = 10^9 kg

be replicated conveniently as needed: you simply measure the wavelength of the standard source to whatever accuracy is required.

Units of length associated with the meter are the centimeter (one hundredth of a meter), millimeter (one thousandth of a meter), and kilometer (1000 meters).

Time. Time intervals are defined by a matching procedure in which the unit of time may be the swing of a pendulum, the emptying of an hourglass, or the completion of some other repeated pattern of motion. The generally accepted standard unit of time is based on the repeating (periodic) motion of the earth around the sun (year) and the rotation of the earth on its axis (day). By means of a pendulum or other such system with a short time of repeating its motion, the second has been defined as 1/86,400 of a mean solar day, which is 1/365.2 ... of a year. As in the case of length, a standard unit of time associated with atomic vibrations has been substituted for the astronomical definition.

Mass. Mass is defined by a matching procedure with an equal-arm balance. The unit of mass could be any object, a stone, or a nail, for example. The accepted unit of mass since 1889 is the kilogram, the mass of a metal cylinder kept under carefully controlled conditions near Paris (Fig. 1.9). The kilogram was intended to be the mass of 1000 cubic centimeters of water at 4° Celsius. Later, more accurate measurements showed that the original determination was slightly in error, so that the reference to water was abandoned. The operational definition of mass makes use of the equal-arm balance (Fig. 1.10), which responds to the downward pull of the earth (commonly known as the weight). Therefore, mass, as we are referring to it here, is called the *gravitational mass*. This idea of mass as intimately connected with the gravitational attraction exerted by the earth will come up again in Section 3.4 where we will explain the related, but distinct, concept of *inertial mass.*

Units of mass derived from the kilogram are the gram (one thousandth of a kilogram), very closely equal to the mass of 1 cubic centimeter of water, and the metric ton (1000 kilograms), very closely equal to the mass of 1 cubic meter of water.

Domain	Size (power of ten)	**Size in meters**	Time (power of ten)	**Time in seconds**
	10^{+24}		10^{+24}	
	10^{+21}	Milky Way (our galaxy)	10^{+21}	
	10^{+18}		10^{+18}	age of Universe
				life on earth
	10^{+15}		10^{+15}	first mammals
		Solar System		first humans
	10^{+12}	orbit of Jupiter	10^{+12}	
Cosmic domain		orbit of Earth		recorded history
	10^{+9}	Sun diameter	10^{+9}	human lifetime
		Earth diameter		year
	10^{+6}	Chicago to NY	10^{+6}	month
				day
	10^{+3}	1 mile	10^{+3}	hour
Macro domain				minute
	10^{0} (1 m.)	Human stride	10^{0} (1 sec.)	heartbeat
	10^{-3}	millimeter	10^{-3}	1 vibration (audible sound wave)
		razor edge		
Micro domain	10^{-6}	single cells	10^{-6}	1 vibration (AM radio wave)
		DNA, proteins		
	10^{-9}	molecules	10^{-9}	1 cycle (1 GHz computer)
		atoms		
	10^{-12}		10^{-12}	1 cycle (1000 GHz computer)
		atomic nuclei		
	10^{-15}		10^{-15}	1 vibration (visible light)
	10^{-18}		10^{-18}	1 vibration (X-ray)
	10^{-21}		10^{-21}	

Figure 1.11 Time and size scale of cosmic, macro, and micro domains.

Other variable factors. It is possible to define units for all other physical variables through definitions based on mass, length, time, and temperature (to be defined in Chapter 10). We will, however, take a different approach, in which we introduce operational definitions for several concepts, such as energy and force, because such an operational procedure makes the physical meaning of the concepts clearer. You will have to accept one disadvantage of this procedure: operational definitions are limited by the technique or operation used and thus will not be the most general ones possible.

Domains of magnitude. We now briefly return to the three domains of magnitude introduced in Section 1.3: the cosmic domain, the macro domain of the everyday world, and the micro domain. Using the definitions of the standard units of measure that we have described, we can approximately characterize the domains by their relationship to these units. Figure 1.11 illustrates this relationship and shows the time and size scale of the various domains.

Summary

"Go, wondrous creature!
Mount where Science guides;
Go measure earth, weigh air, and state the tides;
Instruct the planets in what orbs to run,
Correct old Time, and regulate the Sun."

Alexander Pope
Essay on Man, 1732

The phenomena studied by the physical scientist are highly diverse, ranging from the orbital motion of satellites to the propagation of light, from the turbulent motion of gases in the sun to the structure of the atomic nucleus. The space and time dimensions of phenomena are conveniently divided into three domains: the macro domain, roughly comparable to the human body; the cosmic domain of the very large or very enduring phenomena; and the micro domain of the very small or highly transient phenomena.

In the growth of science, the discovery of new facts and the formulation of new theories go hand in hand. New theories encompass the new facts and may reorganize previously established fields. Working models, thought experiments, and mathematical models are the components of a theory. The terms used to describe models and experiments are related to the real world through operational definitions or to concepts through formal definitions. Measurement (quantitative observation) is introduced through the counting of standard units in the operational definitions of length (distance), time intervals, and gravitational mass.

List of new terms

scientific point of view
scientific "truth"
domains of magnitude:
 micro, macro, cosmic
theory
working model
analogue model
mathematical model
thought experiment
variable factor
constant
formal definition
paraphrase
operational definition
standard object
length: meter
time: year, second
mass: kilogram
equal-arm balance

Problems

1. Give two examples from your own life where you had to revise your expectations (or prejudices) in the light of experience.
2. Describe your feelings toward the study of physics.
3. Describe the values of studying physics as part of a liberal education. Comment on these values from your point of view.
4. Give one or two examples from your own life in which your knowledge of physics was inadequate to the requirements (exclude school experiences).
5. Compare the growth of a city to the growth of science. Does the growth of a city have many similarities to the growth of science? Perhaps new homes correspond to new facts. Perhaps new roads correspond to new theories. Point out similarities and differences. Is the city a good analogue model for science in this respect?
6. Compare the growth of science to various other growth processes. Point out similarities and differences. Are these other examples more or less helpful than the one discussed in Problem 5?
7. Use a dictionary to trace the definition of the word matter. Look up the definition of each major word used to define matter, and so on, until you discover where this process leads. Discuss your discovery and compare it with the approach of this text, which is to leave "matter" as an undefined term (see note in margin on p. 6).
8. Express your preferences with regard to the corpuscular and wave theories of light.
9. Compare scientific "truth" with truth in another domain.
10. Tell which of your senses are most effective in detecting events at the lower limit of the macro domain in space and time. Estimate the magnitude of the smallest length and shortest time interval your senses can detect directly.
11. Tell which of your senses are most effective in detecting events at the upper limit of the macro domain in space and time. Estimate the magnitude of the largest length and longest time interval your senses can detect directly.
12. List examples of indirect evidence (not directly perceived by your sense organs) of phenomena in the macro domain.
13. List examples of direct sensory evidence of phenomena in the micro and cosmic domains. What are some tools used to extend the senses to enable them to cope with phenomena in these domains? Describe the use of these tools and explain whether it leads to direct or indirect evidence.
14. Explain the similarities and differences between a scientific "working model" (such as considering the earth as a uniform, smooth

sphere) and each of the following examples of a "model":
(a) A scale model, such as a model airplane.
(b) A small-scale architectural model of a proposed building.
(c) A model home.
(d) An individual who poses for photographs or paintings, a fashion or artists' model.

15. Carefully examine the system illustrated in Fig. 1.4a.
(a) Propose two (or more) working models that are compatible with all the information given in Fig. 1.4.
(b) Describe one (or more) thought experiments in which your two models exhibit different outcomes. (Such experiments can be used in real experimental tests to eliminate models that lead to a wrong prediction.)

16. Describe two or more working models that apply in an academic field of your choice or in everyday life. For each model, describe some of its properties, how it functions, what observations it explains successfully, and where it fails.

EXAMPLE. Protein-carbohydrate-and-fat model for food. All foods consist of these three materials, in various proportions. The energy (Calorie) value of any food can be found from its content the three materials by a mathematical model. The planning of a balanced diet takes into account the human body's need for the three materials. Gain or loss of weight can be planned on the basis of the Calorie value.

Limitation: it is possible to have a well-balanced diet in terms of proteins, carbohydrates, and fats, yet suffer nutritional deficiencies. The model does not include all the contributions that food makes. Vitamins and minerals are also important, even though they do not contribute to the energy (Calorie) value of food.

Suggested models: computer model for the human brain, gene model for inheritance, "free/efficient market" model for world economy, "economic" model for human beings, demon model for the source of disease.

17. Five blind men investigated an elephant by feeling it with their hands. One felt its tail, one a leg, one a tusk, one an ear, one its side. Describe the analogue models for an elephant they might create individually and by pooling their observations. Describe the implications of this fable for science.

18. Interview three or more children (between ages 7 and 10) to ascertain their ideas as to the source of knowledge and the creation of new knowledge. Ask questions such as, How do we know that 3 + 3 = 6? How do we know that the sun will rise tomorrow? How do we know the earth is round? How do we know how to make a watch (car, rocket, cake....)? Ask questions to probe beyond the first responses. (If possible, undertake this project jointly with several other students so as to obtain a larger collection of responses.) Comment on the responses.

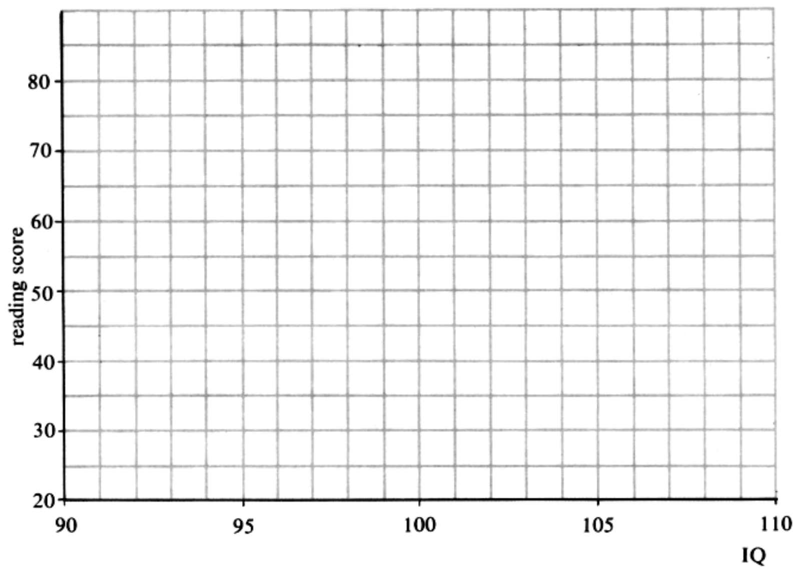


Figure 1.12 Coordinate grid for graph from Problem 19.

TABLE 1-3 READING AND INTELLIGENCE TEST SCORES (PROBLEM 19)

School	*Reading*	*IQ*	*School*	*Reading*	*IQ*
A	33	93	J	51	99
B	59	103	K	31	92
C	57	104	L	51	98
D	46	99	M	69	107
E	48	99	N	73	108
F	54	100	O	48	98
G	52	100	P	75	108
H	52	101	Q	64	105
I	61	103	R	79	111

19. Reading tests and intelligence tests were given sixth graders in a large state. Table 1.3 (above) lists the average scores for schools in eighteen different communities, in order from the largest to the smallest enrollment. Display the data on a graph (Fig 1.12, above), and, if there is a relationship between the two scores, make a mathematical model (in either graphical or algebraic form) for this relationship. Interpret this model. Be careful about making interpretations not actually supported by the given data; explain and criticize whatever assumptions you make, as well as the assumptions that are "hidden" in the data (the test scores).

20. State a formal definition and describe an operational definition for each of the following.
    (a) chair (d) life
    (b) gift (e) person
    (c) teacher (f) scientific literacy
    Comment on the advantages and limitations of the definitions you have constructed.

21. To be constitutional, laws must be applicable to real cases with a minimum of ambiguity. Therefore, they often include operational definitions of the terms that are used in them. Find and report three operational definitions that are part of laws. Discuss the extent to which the inclusion of these operational definitions promotes or restricts the achievement of justice.

22. State three or more operational definitions that you use in your everyday life. The definitions should not deal with profound ideas but may be as simple as: ironing temperature (of a flatiron) is measurable by the "sizzling rate" of a water drop that touches the iron.

23. Write a critique of the hypothesis (beginning of Section 1.2) that the foundations of a person's sense of space and time are laid before birth.

24. Identify one or more explanations or discussions in this chapter that you find inadequate. Describe the general reasons for your judgment (conclusions contradict your ideas, steps in the reasoning have been omitted, words or phrases are meaningless, equations are hard to follow, . . .), and make your criticism as specific as you can.

Bibliography

Introductory physics or physical science textbooks and surveys. These references cover the same material as this text but from a different point of view. Many of them present a more mathematical treatment that uses algebra and trigonometry, but none employs calculus.

K. R. Atkins, *Physics*, Wiley, New York, 1966. Emphasizes interaction point of view, more mathematical.

D. Cassidy, G. Holton, J. Rutherford, *Understanding Physics*, Springer Verlag, 2002. This text, an update of Holton's outstanding 1952 book, is particularly recommended for its historical perspective.

L. N. Cooper, *An Introduction to the Meaning and Structure of Physics*, Harper and Row, New York, 1968. Includes much on quantum physics and relativity, many quotations.

E. R. Huggins, *Physics 1*, W. A. Benjamin, New York, 1968. Presents the interaction point of view. Strong on quantum physics and relativity.

R. B. Lindsay and H. Margenau, *Foundations of Physics*, Wiley, New York, 1936.

V. L. Parsegian, A. S. Meltzer, A. S. Luchins, and K. S. Kinerson, *Introduction to Natural Science, Part One: The Physical Sciences*, Academic Press, New York, 1968. Relates biological, historical, philosophical, sociological, and humanistic material to physical science.

U. Haber-Schaim *et al, PSSC Physics*, Kendall/Hunt, Dubuque, Iowa, 1991. Especially good on wave physics.

E. Rogers, *Physics for the Inquiring Mind*, Princeton University Press, Princeton, New Jersey, 1960. Contains informal explanations, quaint diagrams, and many suggestions for simple experiments.

L. W. Taylor, *Physics, the Pioneer Science*, Dover Publications, New York, 1941. Includes a great deal of historical material.

V. F. Weisskopf, *Knowledge and Wonder: The Natural World as Man Knows It*, 2nd Edition, MIT Press, Cambridge, Mass. 1979. An excellent qualitative lecture series on many facets of physical science. Authored by one of the central players in the discovery of quantum mechanics and quantum electrodynamics, a theorist noted both for his brilliance as well as for his infectious, vibrant personality.

Books and essays on the history and philosophy of science, especially physics:

M. J. Aitkin, *Physics and Archeology*, Wiley (Interscience), New York, 1961.

Francis Bacon. *Novum Organum*, Collier, New York, 1902.

W. I. B. Beveridge, *The Art of Scientific Investigation*, W. W. Norton. New York, 1957.

P. W. Bridgman, *The Logic of Modern Physics*, Macmillan, New York, 1946.

J. Bronowski, *Science and Human Values*, Harper and Row, New York, 1965.

J. Bronowski, *The Common Sense of Science*, Harvard University Press, Cambridge, Massachusetts, 1953.

G. B. Brown, *Science, Its Method and Philosophy*, W. W. Norton, New York, 1950.

H. Butterfield, *The Origins of Modern Science*, Macmillan, New York, 1965.

N. Campbell, *What is Science?,* Dover Publications, New York, 1952.

B. K. Cline, *The Questioners*, Crowell, Collier and Macmillan, New York, 1965.

J. B. Conant, *Science and Common Sense*, Yale University Press, New Haven, Connecticut, 1951.

G. de Santillana, *Origins of Scientific Thought*, University of Chicago Press, Chicago, Illinois, 1961.

A. Einstein and L. Infeld, *Evolution of Physics*, Simon and Schuster, New York, 1938.

A. R. Hall, *The Scientific Revolution 1500-1800*, Beacon Press, Boston, Massachusetts, 1954.

W. Heisenberg, *Physics and Philosophy*, Harper and Row, New York, 1958.

E. H. Hutten, *The Ideas of Physics*, Oliver and Boyd, Edinburgh and London, 1967.

S. L. Jaki, *The Relevance of Physics*, University of Chicago Press, Chicago, Illinois, 1966.

T. S. Kuhn, *Structure of Scientific Revolutions*, University of Chicago Press, Chicago, Illinois, 1962.

S. K. Langer, *Philosophy in a New Key*, Harvard University Press, Cambridge, Massachusetts, 1942.

L. Leprince-Ringuet, *Atoms and Men*, University of Chicago Press, Chicago, Illinois, 1961.

H. Margenau, *The Nature of Physical Reality*, McGraw-Hill, New York, 1950.

E. Nagel, *The Structure of Science*, Harcourt, Brace and World, New York, 1961.

H. Poincaré (S. Gould, Editor), *The Value of Science: Essential Writings of Henri Poincaré* (includes three books: Science and Hypothesis, 1903; The Value of Science, 1905; and Science and Method, 1908), Modern Library, 2001.

J. W. N. Sullivan, *The Limitations of Science*, The New American Library, New York, 1952.

Historical source materials and collections of biographies. These references contain short articles appropriate to many of the later chapters, but these articles are not listed again in the chapter bibliographies. If a topic arouses your interest, you should examine one or more of these references for information supplementary to the text.

A. Beiser, Ed., *The World of Physics*, McGraw-Hill, New York, 1960.

B. Brody and N. Capaldi, Ed., *Science: Men, Methods, Goals*, W. A. Benjamin, New York, 1968.

T. W. Chalmers, *Historic Researches*, Scribner's, New York, 1952.

J. B. Conant and L. K. Nash, *Harvard Case Histories in Experimental Science*, Harvard University Press, Cambridge, Massachusetts, 1957.

L. Hamalian and E. L. Volpe, Ed., *Great Essays by Nobel Prize Winners*, Noonday Press, New York, 1960.

W. S. Knickerbocker, *Classics of Modern Science*, Beacon Press, Boston, Massachusetts, 1962.

H. Lipson, *The Great Experiments in Physics*, Oliver and Boyd, Edinburgh and London, 1968.

O. Lodge, *Pioneers of Science*, Dover Publications, New York, 1926.

W. F. Magie, *A Source Book in Physics*, Harvard University Press, Cambridge, Massachusetts, 1963.

G. Schwartz and P. W. Bishop, *Moments of Discovery*, Basic Books, New York, 1958.

M. H. Shamos, *Great Experiments in Physics*, Holt, Rinehart, and Winston, New York, 1959.

Articles from Scientific American. Some or all of these, plus many others, can be obtained on the Internet at http://www.sciamarchive.org/.

A. V. Astin, "Standards of Measurement" (June 1968).

T. G. R. Brower, "The Visual World of Infants" (December 1966).

D. D. Kosambi, "Scientific Numismatics" (February 1966). A study of the fascinating results yielded by the application of modern science to ancient coins.

H. Zuckerman, "The Sociology of the Nobel Prizes" (November 1967).

The word position has several meanings, two of which, "posture or attitude" and "site or location," are easily confused. We will use the word position always with the latter meaning.

In physics, the terms "relative position" and "relative motion" refer to the fact that position and motion must be defined "relative to" or "in relation to" something other than the object itself.

In common speech, we often simply use ourselves or the earth as the reference object without saying so. For example, we say, "The moon is far away," or "The train is moving." It usually seems unnecessary to say "The moon is far away from me." or "The train is moving with respect to the station." However, in physics we often must describe motion from various reference frames, including especially those in which we are not at the origin and/or in which we are not at rest. Therefore, whenever there is any possibility of confusion, we will explicitly name the reference frame, and you should do the same.

You may associate the word "relativity" with mathematical mystery and scientific complexity, yet the basic concept, which we will try to explain in these pages, is simple. The matters of concern in relativity are the position (location) and motion of objects. The basic concept is that position and motion of an object can only be perceived, described, and recognized with reference to (that is, "relative" to) other objects. When you say, "The physics books are at the left rear of the book store," you refer the position of the books to the entrance and outline of the store. Objects such as the store entrance, to which position or motion are related, are called *reference objects*. Several reference objects used in combination to describe position are said to form a *reference frame (*or *frame of reference)*, and we speak of the position or motion of the original object relative to the reference frame.

If we know the position and motion of an object relative to one reference frame, we might ask about the position and motion relative to a second reference frame. This is the root of the theory of relativity: development of specific mathematical models for relating position and motion as observed relative to one reference frame to position and motion as observed relative to another reference frame. Einstein's theory of relativity is the most complete theory of these relationships. We will describe some aspects of Einstein's work in Section 7.3, but we will not go into the mathematical details in this text.

2.1 Relative position

Look at the girl in the field of daisies (Fig. 2.1). How would you tell someone where she is? Most directly, you could go to the edge of the field and point at her, saying, "The girl is there." By this action, you indicate the position of the girl relative to your outstretched arm and finger.

If you had to describe the girl's position to someone who was not watching, you could say, "She is a little way in from the south edge of the field, near the southeast corner." This statement indicates her position relative to the edges and corners of the field. In other words, it is impossible to describe the position of the girl (or of anything else) without referring to one or more other objects. Even if you were to draw a map of the girl's position, you would have to include on it some objects that could be used to align it with the actual field.

Reference objects and reference frames. For practical purposes, the reference objects must be easy to locate and identify, or they cannot be used as guides in finding the object whose position is being described. It would be hopeless, for example, to try to find the girl in the daisy field if her position were described by saying, "The girl is between two daisy blossoms." Something more distinctive is needed: the edges and comers of the field, as used earlier, or possibly a scarecrow at the center of the field.

The use of reference objects in everyday life is highly varied and adapted to many special circumstances. A piece of furniture, corners of

Figure 2.1 Can you find 3 children hiding among the daises? Describe their relative positions.

a room, street intersections, or a tall building can be used as a reference frame for the complete description of the location of a residence, restaurant, or mailbox.

Examples. An imaginary conversation is recounted in Fig. 2.2. What happened in this conversation? What finally allowed Percy to communicate the location of the hawk without confusion, ambiguity, or absurdity? First, he established a reference frame by selecting a large, easily identifiable branch on the tree and pointing in the direction of the tree. Clyde could grasp this reference frame. Next, Percy specified the direction ("above it") and distance ("the second branch") from the branch to the hawk.

The reference frame first used by Percy consisted of a reference *point* (Percy's body) and a reference *direction* (along Percy's pointing finger). These two components are necessary parts of a reference frame and are defined through more or less easily identified reference objects or earth-based directions, such as north and up. Percy's initial attempt to use the tree as reference point failed because there were several trees.

An overdose of relativity.

Mr. Jones was going to a doctor's office and had never been there before. He called the doctor's office to ask for directions. After the receptionist told him how to get there, he asked whether it was on the north or south side of the street.

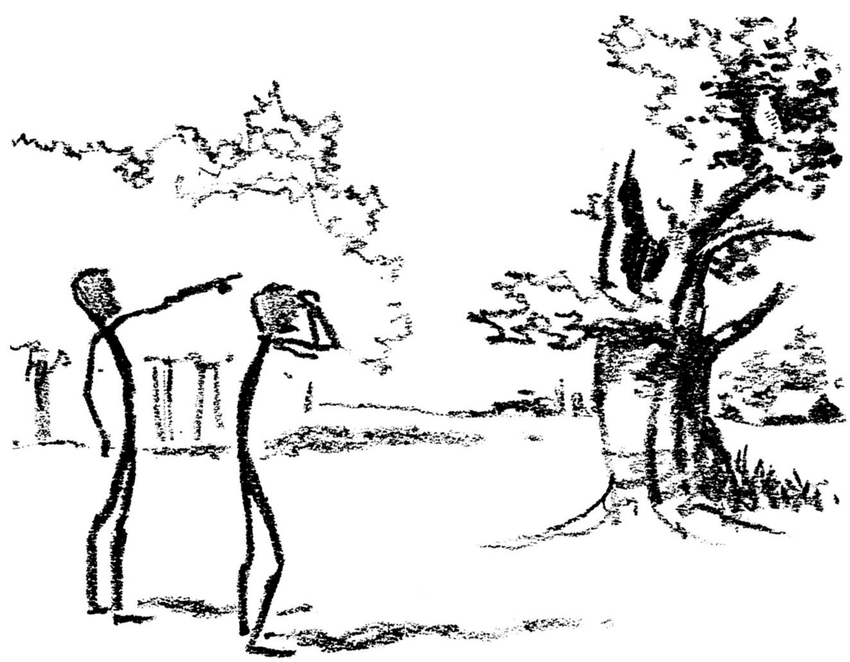
The response: "It depends which way you're walking."

One of the most difficult communications problems is to give instructions for locating a book to a person who is not acquainted with the room in which the book is kept. In such a case it is most helpful to use the person's body as the reference frame by telling him to stand in the door to the room, look for the bookcase on his left, and then scan the middle of the second shelf of that bookcase. This example illustrates how you might use large elements of the environment (the room) to locate smaller ones (the person, and directions defined by the body), and then still smaller ones (the bookcase, ultimately the book) by a narrowing down process.

Describing position is more difficult when you do not have any reference objects to use for a narrowing down process. For example, a passenger on a ship who observes something in the ocean faces this problem. In these circumstances you would have to start with the ship you are on and work outward. You may sight a flying fish "500 yards off the starboard bow" (ahead and to the right), using the ship as reference frame. You could say instead that the fish is "500 yards northwest," using the ship as reference point and compass directions to complete the reference frame.

One-particle model. So far we have been content with describing the position of a very small object that is located at a certain point in space. Real objects, of course, actually occupy an entire region of space, which may be small or large, round or thin, upright or slanted. For a complete description of an object, you therefore should take into account its shape and orientation as well as its location. A useful approach that avoids much unnecessary detail is to make a *one-particle model* for each object of interest. A particle is a very small object

Figure 2.2 Percy and Clyde took a long walk through the county of McDougall.
Percy: Clyde, do you see the falcon sitting in that tree over there?
Clyde: What falcon? In which tree? Where?
Percy: In that big, broken tree over there (pointing his finger).
Clyde: Oh, that tree in front of us! I see it, but I don't see the falcon.
Percy: It's on the branch.
Clyde: There are too many branches. I give up. Let's forget it.
Percy: No, let's start over again. Do you see that broken branch about halfway up the trunk on the right side?
Clyde: Yes, I do.
Percy: Fine. Now look at the second branch above it, on the same side of the tree. Now move to the right and you just have to see the falcon.
Clyde: Oh sure, but that's a hawk.



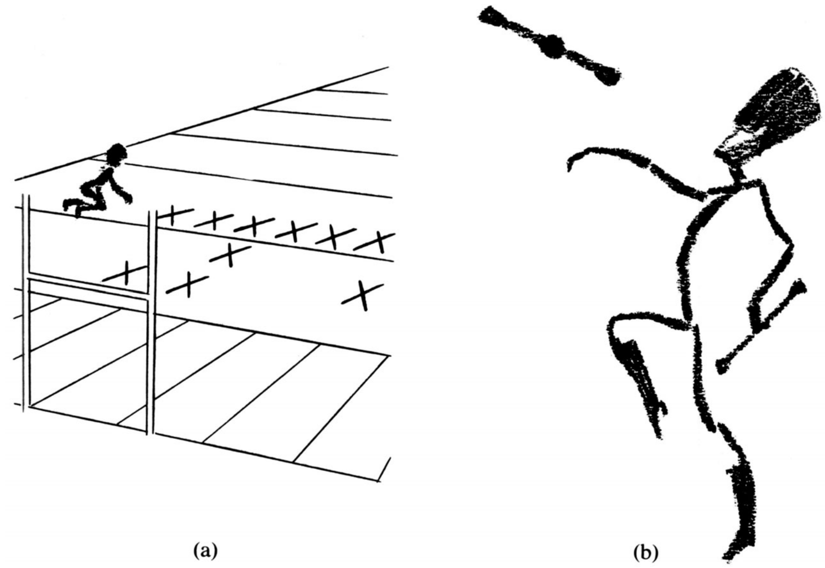


Figure 2.3 Two applications of a one-particle model.
(a) An X represents each football player. Eleven " particles" represent the team.
(b) A working model for the baton is the particle at its center.

that is located at the center or midpoint of the region occupied by the real object (Fig. 2.3). This working model greatly oversimplifies most objects, but is nevertheless accurate enough for most purposes in this text.

Coordinate frames. The laboratory scientist, who tries to describe natural phenomena in a very general way, avoids using incidental objects such as the laboratory walls or table surfaces as reference objects. Instead, a scientist frequently uses a completely artificial reference frame consisting of an arbitrarily chosen reference point and reference direction. The only requirement is to be able to describe the position of objects by using numbers. The numerical measures are called *coordinates;* the reference frame is called a *coordinate frame.* Two coordinate frames in common use are degrees of latitude and longitude, to define position on the earth relative to the equator and the Greenwich meridian, respectively, and distances measured in yards from the end zones and the sidelines on a football gridiron.

Polar coordinates. The procedure of giving the distance from the reference point and the direction relative to the reference direction gives rise to two numbers called *polar coordinates.* The distance may be measured in any unit, most commonly in meters, centimeters, or millimeters. The relative direction is usually measured in angular degrees. How this works is shown in Fig. 2.4. The necessary tools are a ruler to measure distance and a protractor to measure angles. Polar coordinates provide an operational description of the relative position of a point.

A polar coordinate grid, from which you may read the polar

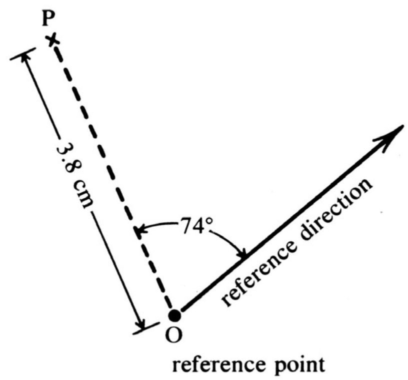


Figure 2.4 Polar coordinates of the point P relative to the reference point 0 and the reference direction indicated by the arrow.

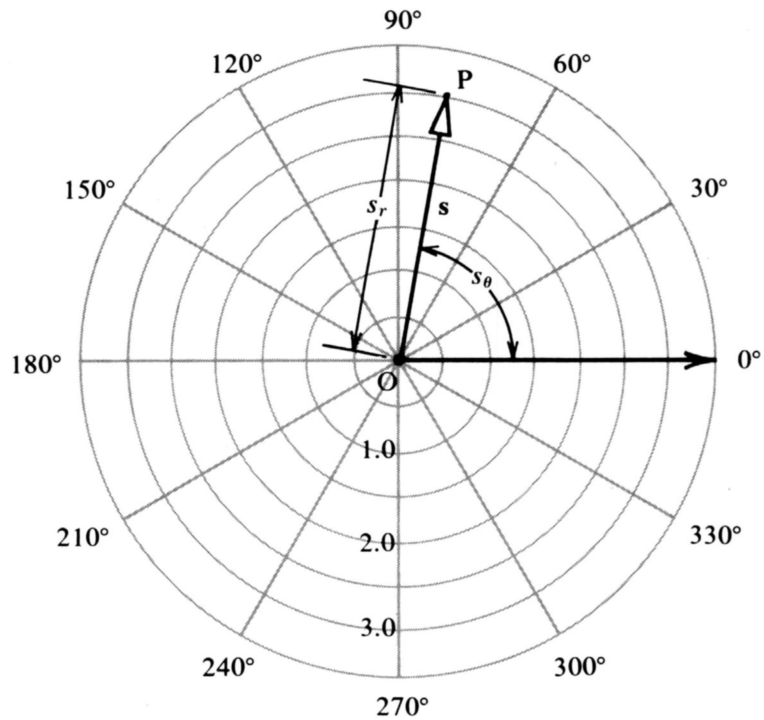


Figure 2.5 Polar coordinates of the point P which has relative position $\mathbf{s} = (s_r, s_\theta) =$ (3 cm, 80°).

s_r and s_θ are examples of variables with subscripts. The subscripts (r and θ) indicate that the main variable (s) has two (or more) distinct values. The subscripts can also be numbers, for example s_1, s_2, s_3 and so on.

You should also be aware of the difference between the meaning of subscripts (described above) and superscripts, for example s^3, which means $s \times s \times s$.

The reference point in polar coordinates is called a pole because the grid lines converge on it, as do the meridians at the poles of the earth.

coordinates of a point directly, is shown in Fig. 2.5. We shall indicate the relative position of the point P by an arrow in the diagram and by the boldface symbol **s** in the text. The polar coordinates will be indicated by the length of the arrow s_r (*r* for radius) and the direction of the arrow s_θ (Greek θ, theta, for angle), as indicated in the figure. Angles are customarily measured counterclockwise from the reference direction.

Examples. An example of the application of polar coordinates is shown in Fig. 2.6. The surveyor is using the direction of the road as the reference direction and the location of his tripod as the reference point. Some of the measurements he has made are given in the figure. These measurements may be used to make a map by locating the objects on a polar coordinate grid, as in Fig. 2.7. Polar coordinates have the intuitive advantage that they represent relative position in the way a person perceives them, namely, with the objects at various distances in various directions from the observer at the center.

Polar coordinates are used to direct aircraft to airports, with the control tower as reference point and north as the reference direction, as well as in other situations where a unique central point exists (e.g., in radar surveillance with the transmitter acting as reference point).

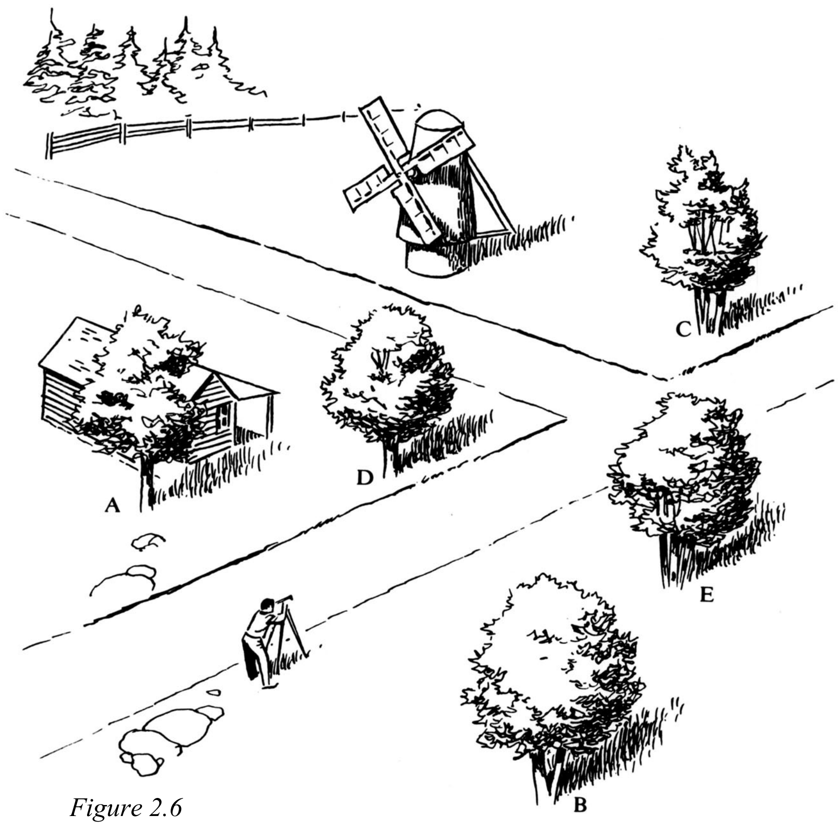


Figure 2.6
Polar coordinates.

	direction	distance
tree A	*80º*	*10 m*
tree B	*270º*	*6 m*
tree C	*10º*	*31 m*
windmill	*40º*	*38 m*

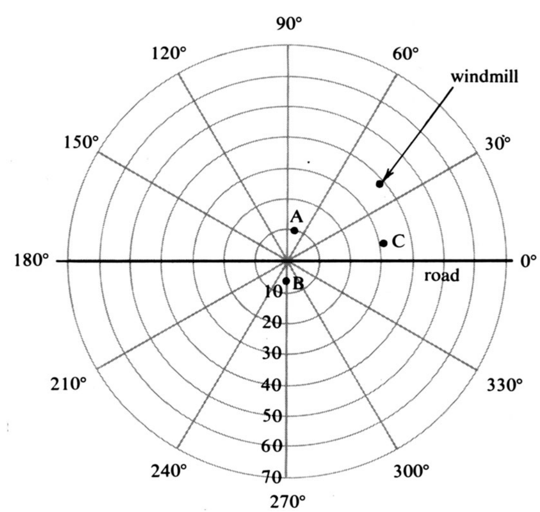


Figure 2.7 The position of the objects in Fig. 2.6 is mapped on a polar coordinate grid.

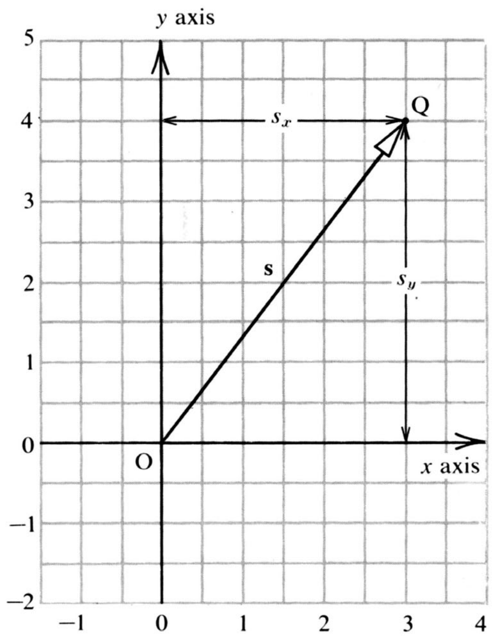


Figure 2.8 Rectangular coordinates s_x and s_y of the point Q, whose relative position is **s** *= $[s_x, s_y]$ = [3.0, 4.0]. The two rectangular axes are indicated by arrows, and the origin of the coordinates by O.*

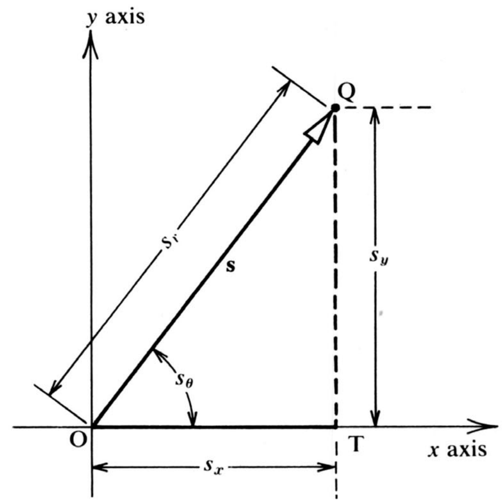


Figure 2.9 Here Fig. 2.8 is redrawn to show the rectangular and polar coordinates of the point Q. These are related by means of the right triangle QTO.

Rectangular coordinates. A particularly useful technique for describing relative position that we will employ extensively later in the course makes use of two lines at right angles to each other called *rectangular coordinate axes* (Fig. 2.8). They are usually labeled the x-axis and y-axis. Their point of intersection is called the *origin of coordinates*. Distances are measured to the desired point Q along lines perpendicular to each axis, and the two measurements obtained are called the *rectangular coordinates* of the point Q relative to the two axes. As before, we introduce the boldface symbol **s** for the relative position of a point and use ordinary letter symbols with subscripts for the rectangular coordinates, this time s_x and s_y. The relative position of point Q is indicated by an arrow from O to Q in Fig. 2.8, just as it was in Fig. 2.5. Sometimes, for the sake of brevity, we will write the rectangular position coordinates in square brackets, with s_x first and s_y second: $[s_x , s_y]$ (Fig. 2.8).

Rectangular coordinates are sometimes called Cartesian coordinates, after René Descartes. Descartes, a French philosopher, first described this coordinate frame in the Discourse on Method*, published in l637.*

You can see that the rectangular coordinates, unlike polar coordinates, do not give directly the distance s_r of a point from the origin. You can find the distance, however, by applying the Pythagorean theorem (Appendix, Eq. A.5) to right triangle QTO in Fig. 2.9, where the distance s_r *is* the length of the hypotenuse OQ:

$$s_r = \sqrt{s_x^2 + s_y^2}\text{, see Ex. 2.1, below.}$$

EXAMPLE 2.1. Relate the polar coordinates of the point Q to its rectangular coordinates (Fig. 2.9) $s_x = 3.0$, $s_y = 4.0$.

Solution:

(a) To find s_r use the Pythagorean theorem (Appendix, Eq. A.5).

$$s_r = OQ = \sqrt{s_x^{\ 2} + s_y^{\ 2}} = \sqrt{(3.0)^2 + (4.0)^2} = \sqrt{25.0} = 5.0$$

(b) To find s_θ, use the definition of the trigonometric ratios (Appendix, Eq. A.6):

$$tangent(s_\theta) = \frac{s_y}{s_x} = \frac{4.0}{3.0} = 1.33$$

$$s_\theta \approx 53^\circ \text{ (from Table A.7)}$$

Rectangular coordinate frames, unlike polar coordinate frames, do not have a single center, as you may observe easily when you compare the polar and rectangular coordinate grids in Fig. 2.5 and 2.8. Whereas there is a unique reference point, the "pole," in the polar grid, it is possible to use any point in a rectangular grid as reference point by selecting the horizontal and vertical lines passing through this point as *x* axis and *y* axis. With R as reference point in Fig. 2.10, for instance, the point Q has the rectangular coordinates [*5, 1*], as you may verify in the figure. Rectangular coordinates, therefore, are advantageous when you are interested in the position of two points or objects relative to one another, rather than only relative to the origin of the coordinate frame. We will use this feature when we calculate changes in the position of a moving object.

2.2 Relative motion

So far we have considered ways of describing the relative position of objects that are stationary. Now, consider objects that change position. When an object changes position, you commonly say that it "moves," or that it is "in motion." But what do you mean by "motion?" Since position is defined relative to a reference frame, it is plausible to expect that motion would also be defined relative to a reference frame. It is therefore customary to use the phrase "relative motion."

Examples of relative motion. Imagine that a truck moving on a roadway is being described relative to two different reference frames. Reference frame A is attached to the roadway, reference frame B to the truck. To make the description simpler and more concrete, we will introduce two observers, one representing each reference frame (Fig. 2.11).

As the truck moves down the road, Observer A reports its position first on his left, then in front of him, then on his right. The position of

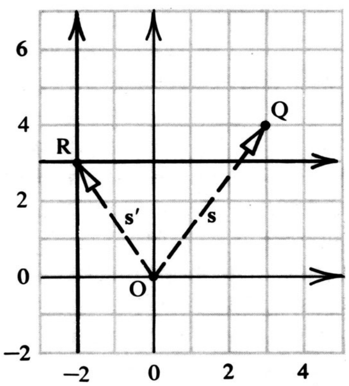
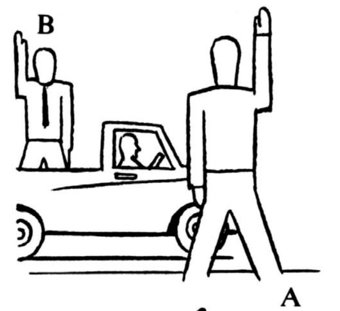
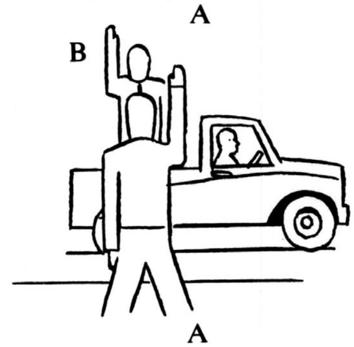


Figure 2.10 (above) The point R is chosen as reference point for describing the relative position of Q. The coordinates of Q relative to R are [5, 1]. (Note the reduced scale of the diagram compared to Figs. 2.8 and 2.9.)





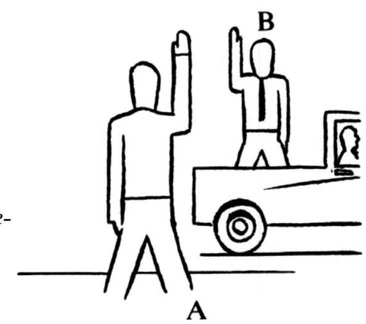


Figure 2.11 The motion of the truck is described relative to the reference frame represented by Observer A and relative to the reference frame represented by Observer B.

the truck relative to Observer A has changed. But Observer B always reports the truck as being in the same position, with the platform under his feet and the driver's cab on his left. The position of the truck relative to Observer B, therefore, has not changed from the beginning to the end of the experiment. Thus, you find that you have two different sets of data, one from each reference frame. In one reference frame you would conclude that the position had changed, and in the other that it had not. If we define relative motion as the change of position relative to a reference frame, then the two observers disagree (but each is correct) not only with regard to the relative position of the truck at various times in the experiment, but also about the truck's relative motion.

We can extend the discussion to motion of other objects. For instance, does the earth move? This depends on the reference frame used to define the earth's position. Relative to a reference frame attached to the earth, to which we are all accustomed, the earth is stationary. Relative to a sun-fixed reference frame, which was introduced by Copernicus and about which you probably studied in school, the earth moves in its orbit. In Chapter 15 we shall describe the resistance which Copernicus and Galileo encountered when they took the sun-fixed reference frame seriously.

Look at some of the consequences of this concept of relative motion. (You may find these consequences fascinating or merely strange, depending on how willing you are to break out of habitual modes of thinking.) In order to determine the position and motion of objects in an experiment, it is important to decide upon a reference frame. You have already noted that observers may disagree about the relative motion of an object. You must also recognize that an observer will always report an object to be stationary in his own reference frame so long as he is attached to that object. Such an observer's report about the motion of the remainder of the world will seem unusual indeed, if the observer's reference frame is attached to a merry-go-round, a satellite in orbit, or even a sewing machine needle.

"For every apparent change in place occurs on account of the movement either of the thing seen or of the spectator, or on account of the necessarily unequal movement of both. For no movement is perceptible relatively to things moved equally in the same directions — I mean relatively to the thing seen and the spectator. Now it is from the Earth that the celestial circuit is beheld and presented to our sight. Therefore, if some movement should belong to the Earth . . . it will appear, in the parts of the universe which are outside, as the same movement but in the opposite direction, as though the things outside were passing over. And the daily revolution . . . is such a movement."

Copernicus
De Revolutionibus
Orbium Coelestium, 1543

Speed and relative motion. Another consequence of the relative motion concept is that different observers might disagree about the direction and the speed of a moving object they both observe. Think about the following example, in which an object is reported to travel at different speeds relative to different reference frames.

The speed of a riverboat going upstream as reported by its passengers looking at the shoreline is a snail-like 1 mile per hour, but the speed as reported by the captain is a respectable 10 miles per hour. Who is right? All steering and propulsion take place in the reference frame of the water. The motion of a riverboat relative to the shore is different from its motion relative to the water unless the water is still. Since the water is flowing downstream at a speed of 9 miles per hour (relative to the shore), and the riverboat is traveling upstream at a speed of 10 miles per hour (relative to the water), the speed of the riverboat (relative to the shore) is 1 mile per hour. Thus, the conflicting reports of the two observers are understandable and correct, for they are observing from different reference frames. The question, "Who is right?" can only be answered, "Each is right from his own point of view."

Here is a second example, in which we would like you to imagine you are each of the observers in turn:

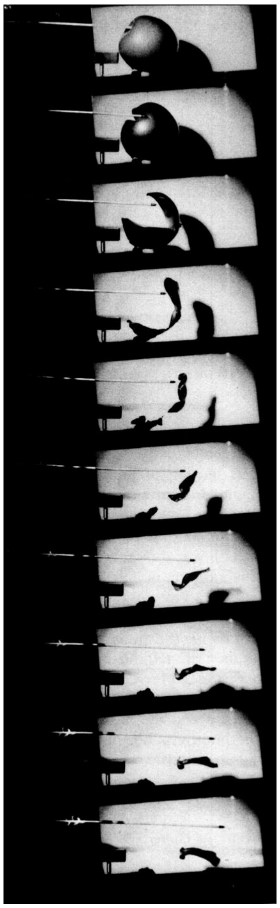
Three cars, A, B, and C, are traveling north on a highway at speeds of 55, 65, and 75 miles per hour, respectively. Observers attached to each car make the following reports.

What does the observer in car B report about the speed of the roadside? What does the observer in car C report about the speed of car A and the roadside?

Observer in car A: Relative to me, car B is traveling north at 10 miles per hour, car C is traveling north at 20 miles per hour, car A (my car) is stationary, and the roadside is traveling south at 55 miles per hour.

Observer in car B: Relative to me, car A is traveling south at 10 miles per hour and car C is traveling north at 10 miles per hour.

Observer in car C: Relative to me, car B is traveling south at 10 miles per hour, and car C is stationary.



Recording and reproducing relative motion. Since relative motion is a transitory phenomenon, it cannot be recorded on a diagram or map with the ease with which relative position can be recorded. The motion-picture film is the most familiar way of recording and recreating relative motion.

Motion pictures. Motion pictures consist of a strip of photographs (Fig. 2.12) that show a scene at very short intervals (approximately 1/24 second). Of course, the scene changes, but does not change much in this short time. When the pictures are rapidly projected in the correct order, the viewer's eye and mind perceive smooth motion. If the pictures are taken with too great a time interval, so that the scene changes significantly between, then the smooth motion becomes jerky.

Flip books. Another way to represent relative motion is through flip books. Flip books create the illusion of motion through the same device as motion pictures, a series of pictures that show the same scene with slight changes in appearance. The pictures are bound in a book and are viewed when the pages of the book are flipped. With a flip book, you can examine each individual scene more easily than with a motion picture, you can control the speed, and you can view the sequence both "forward" and "backward" by starting at the front or the back of the book.

Multiple photographs. Still another technique for representing but not recreating motion is the multiple photograph. This is produced by taking many pictures at equal short time intervals on the same piece of film as in the example shown in Fig. 2.13. The multiple photograph gives a record of the path of the racquet and of the ball during a serve.

Blurred photographs. Even a single photograph can give evidence of motion when the camera shutter remains open long enough for the image projected onto the photographic film to change appreciably.

Figure 2.12 A section of motion-picture film taken at high speed, showing an arrow bursting a balloon. Harold Edgerton, the master of high speed photography, took the photograph by means of a rotating prism synchronized with a strobe [regularly flashing] light. In 1940, Edgerton won an Academy Award for movies made with this type of camera. For other photos by Edgerton see Stopping Time *by G. Kayafas, listed in the bibliography at the end of this chapter.*

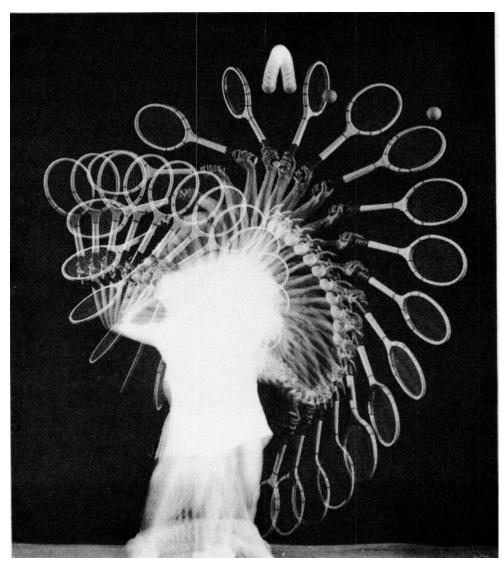


Figure 2.13 Multiple exposure photograph of a tennis serve, taken by Harold Edgerton. Can you estimate the time interval between exposures?

Figure 2.14 How many people are shown in this picture?

This can happen either because the photographic subject moves (Fig. 2.14), or, as every photographer knows, because the camera moves. In other words, the significant fact is motion of the subject relative to the camera. Indeed, an experienced photographer can "stop" the motion of a racing automobile by purposely sweeping his camera along with the automobile (Fig. 2.15). The automobile appears sharply in this picture, but the background, which was moving relative to the automobile and therefore relative to the camera, is blurred.

Example. An illustration of how significant relative motion can be was discovered by Berkeley physicist Luis W. Alvarez in a magazine reproduction of part of a motion-picture film showing the assassination of President John F. Kennedy. The President was riding in a motorcade, and Alvarez noticed something exciting in photograph 227: the motorcade in the photo was blurred, but the background and foreground were sharp. This was in contrast to most other photos, where the background was blurred and the motorcade was sharp. Apparently, Alvarez reasoned, the photographer had been sweeping his camera sideways to keep it lined up with the moving car but had suddenly stopped the

Figure 2.15 The racing car is stationary relative to the camera. Were its wheels stationary?

motion for a fraction of a second. The film, which showed the relative motion of camera and photographed objects by a steady change in picture from frame to frame, showed a change in this relative motion, a change that Alvarez ascribed to the photographer's neuromuscular reaction to the sound of a rifle shot. Further investigation of the original film in the National Archives, and of human flinching reactions to sudden sounds, confirmed Alvarez's interpretation of the blurred motorcade as evidence that a rifle was fired at that instant.

FORMAL DEFINITION
The average speed is equal to the ratio of the distance traversed divided by the time interval required to traverse the distance.

Equation 2.1

distance traversed = Δs
(pronounced "delta ess")
time interval = Δt
(prounced "delta tee")
average speed = v_{av}

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

*Note: The "*Δs*" and "*Δt*" symbols stand for single quantities and <u>do not</u> indicate multiplication of* Δ *by "*s*" or by "*t*". The meaning of the* Δ *symbol will be explained further below (Section 2.3).*

Units of speed:
meters per second (m/sec)
miles per hour (mph)
feet per second (ft/sec)

1 m/sec ≈ *2.2 mph* ≈ *3.3 ft/sec*

1 mph ≈ *0.45 m/sec* ≈ *1.5 ft/sec*

(≈ *indicates approximately equal)*

Definition of speed. Up to this point we have used the word "speed" without defining it. Automobile speed in miles per hour is usually read directly off the dial of a speedometer, the speed of a runner is expressed in his or her times for a particular distance, wind speed is indicated by a device called an anemometer, and so on. To be comparable with one another, these speeds must all be derived from the same definition. As the unit of speed (miles per hour) suggests, the generally accepted definition (at left) is the rate at which distance is traversed (Eq. 2.1).

This definition can be applied whenever a distance and a time measurement have been made. You are probably familiar with the highway "speedometer checks," roadside markers that identify a one-mile stretch; by driving at a steady speed and observing the time required to traverse the distance, you can calculate the speed (Example 2.2).

EXAMPLE 2.2. Find the speed if the distance traversed and the time interval are given.

(a) $\Delta s = 2.5\,m \qquad \Delta t = 4.5\,\text{sec}$

$$v_{av} = \frac{\Delta s}{\Delta t} = \frac{2.5\ m}{4.5\ sec} \approx 0.56\ m/sec$$

(b) $\Delta s = 140\,m \qquad \Delta t = 0.4\,\text{sec}$

$$v_{av} = \frac{\Delta s}{\Delta t} = \frac{140\ m}{0.4\ sec} \approx 350\ m/sec$$

(C) $\Delta s = 1\,mile \qquad \Delta t = 4\,\text{min}$

$$v_{av} = \frac{\Delta s}{\Delta t} = \frac{1\ mile}{4\ min} \approx 0.25\ mile/min$$

$$= 15\ mph \approx (15 \times 0.45)\ m/sec = 6.8\ m/sec$$

The definition can be applied in many other circumstances, too, where the distance and time interval may be measured in any convenient unit. The speed of a taxicab in a large city, for instance, may be described as six blocks per minute, the speed of a bus may be only three blocks per minute, the speed of an elevator may be one floor in three seconds (one-third floor per second or 20 floors per minute), and so on.

You can convert from one to another of these units of speed if you know how the units of distance and time compare (number of feet per floor, number of seconds per minute).

"If two particles are carried at a uniform rate, the ratio of their speeds will be the product of the ratio of the distances traversed by the inverse ratio of the time intervals occupied."
Galileo Galilei
Dialogues Concerning Two New Sciences, 1638

Curiously, the work of Galileo, who was the first to investigate moving bodies systematically and quantitatively, contains no reference to this idea of speed as a numerical quantity (v) equal to the ratio of distance divided by time. Instead, he always compared two or more speeds with one another (often by comparing the times to go equal distances, or by comparing the distances traveled in equal times), and he was able to derive and state his results by using ratios of distances (or times) to one another. One of Galileo's major contributions was a clear understanding of what we now call "average speed" and "instantaneous speed." We now explain these two key concepts.

Average speed. When you think of a bus making its way in city traffic, you immediately realize that the speedometer reading has little direct connection with a measured speed of, say, three blocks per minute. After all, the bus is stopped a good fraction of the available time. The speedometer needle may swing from 0 miles per hour (the bus is stopped) up to 20 or even 30 miles per hour while the bus is moving, and then back to 0 miles per hour again at the next stop. If you count how many blocks the bus travels in a minute, you include the stops and the motion. The speed determined in this way is called the *average speed*, because it is an average value intermediate between the maximum and minimum values. The average speed is always referred to a certain distance or time interval, such as the average speed over a mile of highway (speedometer check) or in a minute of city driving (bus and taxi examples). A car that required a minute to drive 1 mile on the highway was traveling at the average speed of 1 mile per minute, 60 miles per hour, or about 90 feet per second.

FORMAL DEFINITION
The instantaneous speed is equal to the average speed measured during an "instant." An instant is a time interval short enough that the speed does not change to a significant extent.

Instantaneous speed. After this explanation, you may wonder what the car's speedometer indicates. The speedometer indicates the "actual speed" of the car. The actual speed is equal to the average speed if the car is driven steadily without speeding up or slowing down. In this way the speedometer check can be used as intended by the highway builders. In other words, the average speed, which can be measured in the standard units of distance and time, is used to calibrate the speedometer dial.

OPERATIONAL DEFINITION
The instantaneous (or actual) speed is equal to the number shown by a speedometer.

There is a second relation between average speed and actual speed – a relation that has led to the term *instantaneous speed* for the latter. Imagine the average speed measured during a very short time interval, such as 1 second or less. During such a short time interval, the car has barely any possibility of speeding up or slowing down. Hence the average speed in this short time interval is practically equal to the actual speed. Since a very short time interval is called an instant, the name instantaneous speed is generally used (Eq. 2.2).

Equation 2.2

instantaneous speed = v

$$v = \frac{\Delta s}{\Delta t}$$

(Δt *is an instant, an interval of time chosen short enough so that the speed does not change to a significant extent.)*

How short is an "instant"? The instant is defined to be so short that the speed of the moving object does not change appreciably. Just how short it must be depends on the motion that is being studied. For a car that accelerates from a standing start to 60 miles per hour in 10 seconds, the instant must be considerably shorter than 1 second. For a bullet being fired, an instant must be very much shorter yet, for the entire time interval during which the bullet accelerates inside the gun barrel is much,

much shorter than 1 second. At the other extreme, consider the ice in a glacier slowly gliding down a mountain valley. For this motion, even a day may be a brief instant because years elapse before the speed changes.

Applications. Since the average speed is defined by means of a mathematical formula ($\Delta s/\Delta t$), you can use mathematical reasoning (Section A.2) to solve a variety of problems. For instance, you can compute the distance traversed by a moving object if you know its average speed and the travel time (Section 1.3, Eq. 1.2, and Fig. 1.6). Or you can compute the time required for a trip if you know the average speed and the distance to be covered. These ideas are illustrated in Example 2.3.

EXAMPLE 2.3

(a) Find the distance if the average speed and time interval are given. How far does a pedestrian walk in 1.6 hours?

Solution: We wish to use Equation 2.1 to find Δs; thus we multiply both sides of Equation 2.1 by Δt to get: $\Delta s = v_{av}\, \Delta t$

For a pedestrian we can estimate $v_{av} \approx 3$ mph and $\Delta t = 1.6$ hours.
Thus $\Delta s = v_{av}\, \Delta t = 3$ mph x 1.6 hours ≈ 5 miles.

(b) Find the time interval required if the average speed and distance are given. If a bullet's average speed is 700 m/sec, how long does a bullet take to travel 2000 meters?

Solution: We wish to find Δt; thus we multiply both sides of Equation 2.1 by Δt and divide by v_{av} to get: $\Delta t = \dfrac{\Delta s}{v_{av}}$

For the bullet,

$$v_{av} = 700\, m/sec = 7\times 10^2\ m/sec$$
$$\Delta s = 2000\, m = 2\times 10^3\ m$$
$$\Delta t = \frac{\Delta s}{v_{av}} = \frac{2\times 10^3\ m}{7\times 10^2\ m/sec} \approx 0.28\times 10 = 2.8\, sec.$$

2.3 Displacement

The concept connecting relative position with relative motion is the change of relative position, which enables you to apply coordinate techniques to motion. For example, when a particle moves from one point R to another point Q (Fig. 2.16), then its position relative to any reference point fixed on the page changes. The change in position of a moving object is called the *displacement* because you can think of the moving object being displaced from one point to the other, from R to Q. By marking the successive displacements of a moving object, you can trace its path in space (Fig. 2.17).

The symbol for displacement is the boldface $\mathbf{\Delta s}$ with the Greek $\mathbf{\Delta}$

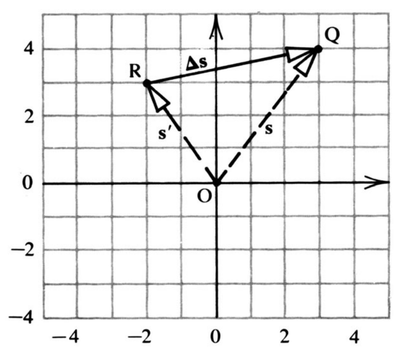


Figure 2.16 *The dashed arrows represent the positions of points Q and R relative to the coordinate axes. The solid arrow represents the displacement* **Δs** from *R to Q.*

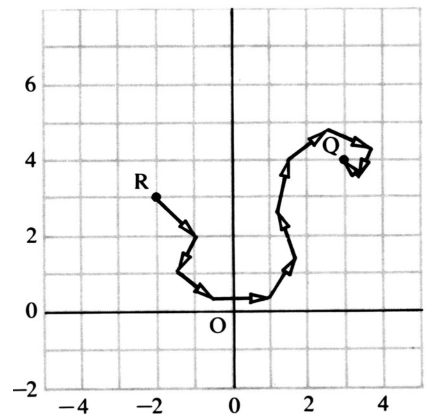


Figure 2.17 Arrows represent the successive small displacements of a particle whose overall (or net) displacement is from R to Q.

Equation 2.3

Relative position of Q

$$s=[3,4]\begin{cases} s_x=3 \\ s_y=4 \end{cases}$$

Relative position of R

$$s'=[-2,3]\begin{cases} s_x'=-2 \\ s_y'=3 \end{cases}$$

Equation 2.4

Change of relative position from R *to* Q *(displacement* **Δs**)

$$\begin{aligned} \Delta s_x &= s_x - s_x' \\ &= 3-(-2)=5 \\ \Delta s_y &= s_y - s_y' \\ &= 4-3 = 1 \\ \mathbf{\Delta s} &= \mathbf{s - s'} \\ &= [s_x - s_x', s_y - s_y'] \\ &= [5, 1] \end{aligned}$$

Note that the x- and y- components of **Δs** *are independent of one another and are calculated separately.*

(delta) standing for "difference" and **s** standing for relative position. The boldface symbol **Δs** is a single entity and does not signify multiplication of two factors. It should be distinguished from the symbol Δs, the distance traveled, which was used in the definition of speed (Eq. 2.1). The displacement (**Δs)** is a more complex quantity than simply the distance traveled (Δs). The displacement **Δs** includes the distance traveled (Δs) **and** the direction of that movement in space.

With the help of the coordinate grid in Fig. 2.16, you can find the coordinates of Q and R relative to the origin O (Eq. 2.3). The coordinates of the change in relative position, which are called the components of the displacement from R to Q, are written Δs_x and Δs_y. They are equal to the differences of the coordinates of Q and of R relative to O (Eq. 2.4).

In a diagram, a position or displacement will be indicated by an arrow (with an open arrowhead, Figs. 2.16 and 2.17) from the reference or starting point at the tail to the actual or final point at the head. The length of the arrow represents the magnitude of the displacement, and the direction of the arrow represents the direction of the displacement. If several arrows have to be drawn, then their tails, magnitudes, and directions must be properly related. As you will discover in later chapters, the arrow description of a magnitude and a direction in space will be used for force, velocity, and other physical quantities, as well as for displacements

Placement of arrows. Consider, for instance, the surveyor (Fig. 2.6), whose measurements were represented by relative position arrows in Fig. 2.7. In this diagram the tails of all the arrows would be placed at the same point, which represents the surveyor's benchmark. If, however, you want to track a sailboat moving on a zigzag course

(Fig. 2-18), then you must represent the displacement on each straight part by an arrow whose tail is placed at the head of the arrow representing the preceding displacement. Thus, you obtain a map of the boat's path, and you can find the overall displacement from the starting point to the finish point by interpreting the map.

Addition of displacements. The process of combining the displacements to find the overall displacement by placing the tail of one arrow at the head of the previous one is called *addition of displacements*, and the overall displacement is called the *sum.* This is analogous to the

Figure 2.18 The sailboat proceeds from the starting point to the finishing point along a six-part zigzag course. The individual displacements $\mathbf{\Delta s_1}$, $\mathbf{\Delta s_2}$, $\mathbf{\Delta s_3}$, $\mathbf{\Delta s_4}$, $\mathbf{\Delta s_5}$, *and* $\mathbf{\Delta s_6}$ *combine (add together) to give the sum, or overall displacement* $\mathbf{\Delta s}$. *(Scale: length of side of small square in graph below = 1/2 mile.)*
First, we find the various displacements by counting squares on the figure below:

$\mathbf{\Delta s_1}$ = [*+2.0, +3.0*] $\mathbf{\Delta s_3}$ = [*+5.0, +3.0*] $\mathbf{\Delta s_5}$ = [*+2.5, -1.5*]

$\mathbf{\Delta s_2}$ = [*+1.0, -3.0*] $\mathbf{\Delta s_4}$ = [*-0.5, -2.0*] $\mathbf{\Delta s_6}$ = [*-0.5, +5.5*]

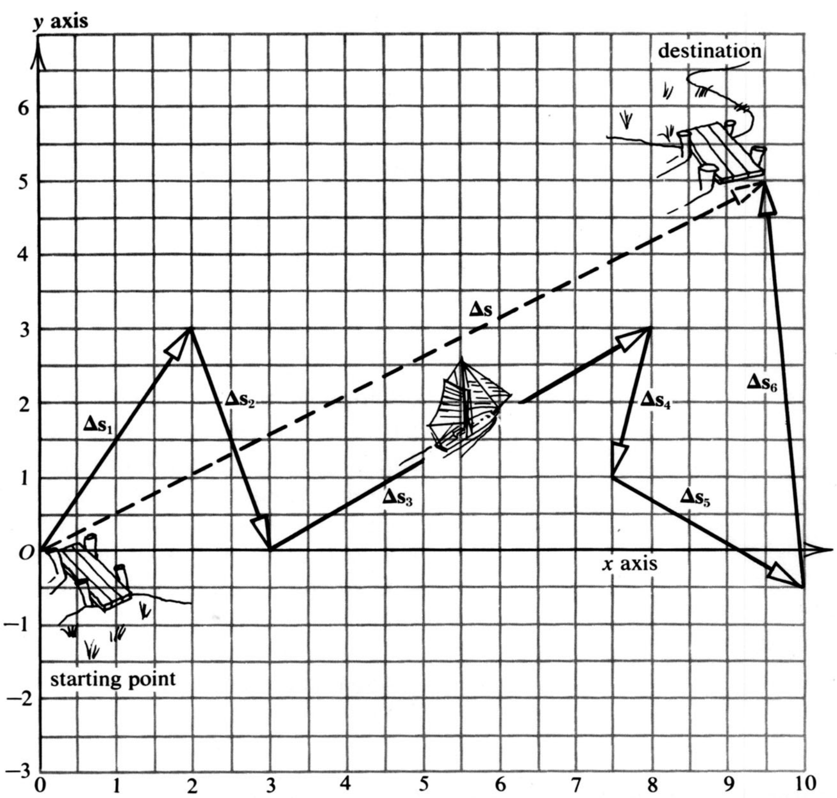
$\mathbf{\Delta s} = [\Delta s_x, \Delta s_y]$ = [*+9.5, +5.0*] (counted directly from figure below)

We can check the above result by adding the individual displacements to find the sum: $\mathbf{\Delta s_1 + \Delta s_2 + \Delta s_3 + \Delta s_4 + \Delta s_5 + \Delta s_6 = \Delta s}$

Now, to actually find the x- and y-components of $\mathbf{\Delta s}$***, we must add the individual x- and y-components separately, so they do not get mixed up with one another!***

x components : $\Delta s_x = 2.0 + 1.0 + 5.0 - 0.5 + 2.5 - 0.5 = 9.5$ *(agrees with above)*

y components: $\Delta s_y = 3.0 - 3.0 + 3.0 - 2.0 - 1.5 + 5.5 = 5.0$ *(agrees with above)*



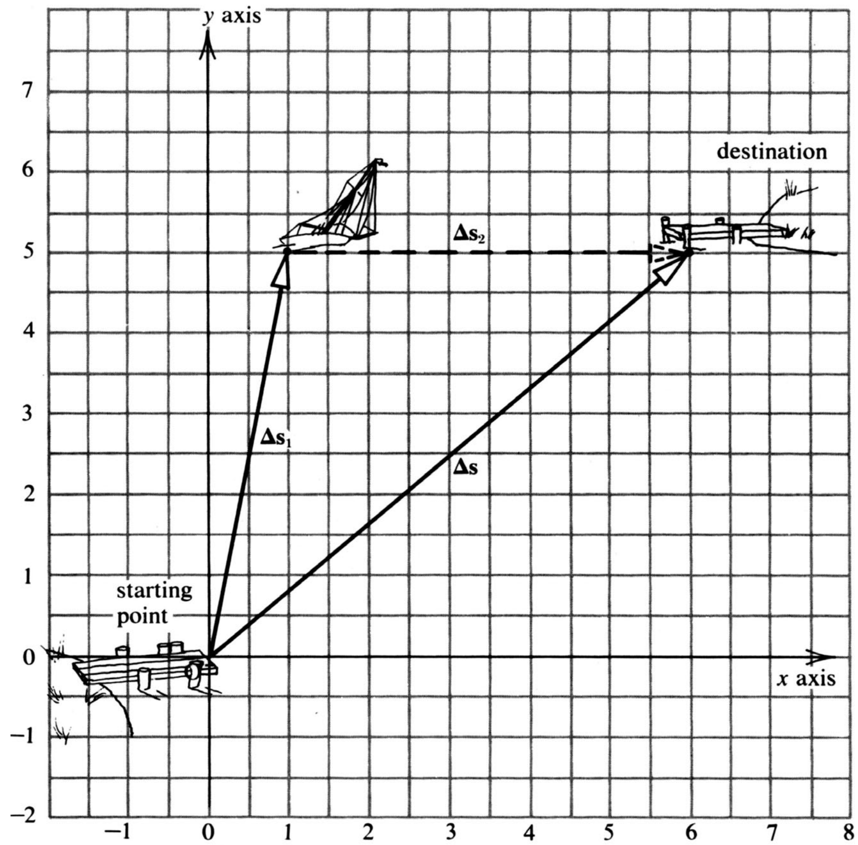


Figure 2.19 The sailboat has accomplished displacement $\mathbf{\Delta s_1}$ *and still needs to make displacement* $\mathbf{\Delta s_2}$ *to reach its destination. The displacement* $\mathbf{\Delta s_2}$ *is the difference between* $\mathbf{\Delta s}$ *and* $\mathbf{\Delta s_1}$*. (Scale: 1 square = 1/2 mile.)*
$\mathbf{\Delta s_1}$ *= [1, 5],* $\mathbf{\Delta s_2}$ *= [5, 0],* $\mathbf{\Delta s}$ *= [6, 5].*
$\mathbf{\Delta s_2} = \mathbf{\Delta s} - \mathbf{\Delta s_1}$ = *[6, 5]* - *[1, 5]* = [6 - 1, 5 - 5] = [5, 0]. (**Note that the x- and y-components are subtracted separately so they do not get mixed together!**)
As you can see in the figure, $\mathbf{\Delta s_2}$ indeed has an x-component of 5 and a y-component of 0.

Finding the sum (or difference) of displacements*: the x- and y-components are independent and <u>must be kept track of separately</u>; thus x-components are only added to (or subtracted from) other x-components, and y-components are only added to (or subtracted from) other y-components.*

addition of numbers, where \$100 combined with \$60 gives the sum of \$160. The graphical process of adding displacements is illustrated in Fig. 2.18. You can also use normal arithmetic to do this if you know the rectangular components of each displacement. The rectangular components may be read off Fig. 2.18 and are listed in the legend to that figure. It is clear that the *x* component of the overall displacement is the sum of the *x* components of the individual components, and the same is true of the y components. It is essential to keep the x- and y-components separate.

Subtraction of displacements. The course of the sailboat in Fig. 2.18 gave a natural illustration of the sum of displacements. To find illustrations of the difference of displacements, consider first two ways of interpreting the difference of two numbers: what is left over after part is removed and what is needed to obtain a larger quantity. The first way applies when you have \$100 and spend \$60; you are left with the difference, which is \$40. The second way applies when you want \$100 and have \$60; you still need the difference, which is \$40. The second interpretation can be applied to displacements. If you are in the sailboat, are aiming for a destination at a displacement $\Delta\mathbf{s}$ from the starting point, but have only made the progress described by the displacement $\Delta\mathbf{s}_1$, the displacement $\Delta\mathbf{s}_2$ must still be traversed (Fig. 2.19). The displacement $\Delta\mathbf{s}_2$ is the difference between the goal $\Delta\mathbf{s}$ and the partial achievement $\Delta\mathbf{s}_1$, that is, $\Delta\mathbf{s}_2 = \Delta\mathbf{s} - \Delta\mathbf{s}_1$. The difference may be found either graphically or arithmetically from the rectangular displacement components by subtraction as shown in Fig. 2.19.

Equation 2.5

$$\frac{\Delta\mathbf{s}}{b} = \left[\frac{\Delta s_x}{b}, \frac{\Delta s_y}{b}\right]$$

EXAMPLE

$\Delta\mathbf{s}$ *= [14, 5] and b = 4*

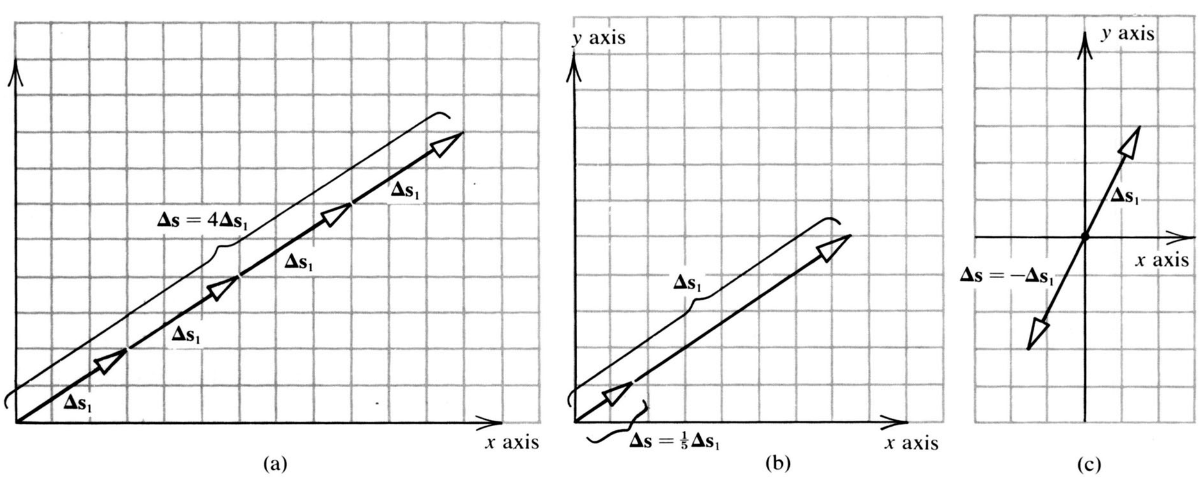
Then

$$\frac{\Delta\mathbf{s}}{b} = \frac{[14,5]}{4} = [3.5, 1.25]$$

Multiplication and division. Certain other arithmetic operations can be carried out with displacements by performing these operations on all the rectangular components of the displacements, just as you have calculated sums and differences by applying the appropriate arithmetic operation to the rectangular components. By adding a displacement to itself repeatedly (Fig. 2.20a), you obtain a multiple of the displacement. You can also divide a displacement into equal parts, such that each part is a fraction of the original displacement (Fig. 2.20b). Finally, you can find the negative of a displacement, which is just a displacement of equal magnitude and in the opposite direction from the original displacement (Fig. 2.20c).

An important algebraic concept is the division of a displacement by a number (Eq. 2.5), which will be used in the definition of velocity and

Figure 2.20 Arithmetic operations with displacements.
(a) Multiple displacements.
(b) Fractional displacement.
(c) Negative displacement.



acceleration in Chapter 13. Specifically, a displacement divided by a number is just another, smaller displacement in the same direction; to calculate this, simply divide each rectangular component by the number, as illustrated in Eq. 2.5.

Summary

Position and motion of objects can only be observed and described relative to reference frames. Position or motion of an object may differ when described relative to different reference frames. A reference frame may be centered on you or on any other point in space that is stationary or in motion relative to you. Quantitative ways of describing relative position make use of polar coordinates (s_r , s_θ) and rectangular coordinates [s_x, s_y]. A quantitative description of relative motion makes use of the average speed, which is defined as the distance traveled divided by the time interval required (Eq. 2.1).

The change of an object's relative position is called the displacement. The displacement has a magnitude and a direction in space. It is represented in diagrams by an arrow and is described quantitatively by polar coordinates or rectangular components, [Δs_x , Δs_y]. The processes of arithmetic (addition, subtraction, multiplication and division) can be carried out with displacements in rectangular coordinates by calculating the x- and y-components separately.

List of new terms

reference object	coordinate frame	displacement
reference frame	polar coordinates	component
reference point	rectangular coordinates	average speed
reference direction	coordinate axis	instantaneous speed
particle	origin of coordinates	

List of Symbols

s (bold)	relative position in space		
[s_x , s_y]	rectangular position coordinates	Δs	distance traversed
		Δt	time interval
[s_r , s_θ]	polar position coordinates	v	speed
Δs (bold)	displacement		
[Δs_x, Δs_y]	displacement components (or displacement rectangular coordinates)		

Problems

1. Lie on your bed on your back. Use your head as reference point and the directions front, back, right, left, above, below (or combinations of these) to describe the position of objects relative to this reference point. Give the approximate direction and distance of several of the following: pillow, lamp, radio, door, your feet, and so on.

2. Describe the position of your bedroom by using the building as reference frame. (Do *not* give detailed instructions as to how a person could walk to your bedroom.)

3. Describe the position of the children in Fig. 2.1 by using the picture edge as reference frame.

4. Find an alternate way of locating the hawk for Clyde (Fig. 2.2).

5. (a) Estimate the polar coordinates of the house and trees D and E in Fig. 2.6.
(b) Mark the location of the house and trees D and E on the map in Fig. 2.7.

6. Find the distance from R to Q in Fig. 2.10.

7. Find an arithmetic relationship among the coordinates of Q relative to O, R relative to O, and Q relative to R in Fig. 2.10.

8. A man on the earth travels 10 miles south, then 10 miles east, then 10 miles north. After the 30-mile trip is finished, he is back at his starting point. Identify his starting point.

9. Give two examples from everyday experience where you intuitively identify motion relative to a reference frame that is moving relative to the earth.

10. Explain how a motion-picture strip can be used to determine the speed of a moving object. Refer to average and instantaneous speeds. Apply your method to Fig. 2.12. You should estimate the approximate distances in the figure.

11. Explain how a multi-flash photograph like Fig. 2.13 can be used to determine the speed of the moving object at various points along the path. Apply your method to an object in Fig. 2.13. You should estimate the approximate distances in the figure.

12. Measure the average speeds of two or three objects in everyday life. You may use any convenient units, but you should convert to standard units (meters per second). Explain for each example how you chose to define "average."

Problems 13-16 ask you to compare or describe motion. You should describe the path(s) of the motion in words, name the reference frame(s) and make a drawing(s). You should also describe the speed of the motion.

13. A record is being played on a phonograph. Compare the motion of the needle relative to the phonograph base with the needle's motion relative to the record on the turntable.

14. Describe the motion of the moon relative to a reference frame fixed on the sun. Use a one-particle model for the moon.

15. Describe the motion of the earth relative to a reference frame fixed on the surface of the moon. Use a one-particle model for the earth.

16. Describe the motion of the sun relative to a reference frame fixed on the moon. Use a one-particle model for the sun.

17. A girl is pedaling a bicycle in a straight line.
(a) Describe the motion of the tire valve relative to the axle of the wheel.
(b) Describe the motion of the tire valve relative to the road.
(c) Describe the motion of one pedal relative to the road.
(d) Describe the motion of one pedal relative to the other pedal.

18. Using the two different methods outlined below, find the displacement from tree A to the windmill in Fig. 2.7.
(a) Use ruler and protractor to solve the problem geometrically. State the result in polar coordinates. You answer should have both a distance (s_r) and an angle (s_θ), expressed as [s_r , s_θ].
(b) Impose a rectangular coordinate frame (graph paper) on Fig. 2.7 and solve the problem arithmetically. You answer should have both an x-component (s_x) and a y-component (s_y), expressed as [s_x , s_y].

19. These questions deal with Fig. 2.18.
(a) Find the combined displacement of the second and third legs of the sailboat's course.
(b) Find the combined displacement of the fourth, fifth, and sixth legs of the sailboat's course.
(c) Find the combined displacement of the first, third, and fifth legs of the course.
(d) Find the displacement still required for the sailboat to reach its destination after the first leg of the course.
(e) Find the displacement still required for the sailboat to reach its destination after the fourth leg of the course.

20. Identify one or more explanations or discussions in this chapter that you find inadequate. Describe the general reasons for your dissatisfaction (conclusions contradict your ideas, or steps in the reasoning have been omitted; words or phrases are meaningless; equations are hard to follow; etc.) and pinpoint your criticism as well as you can.

Bibliography

G. Galilei, *Dialogues Concerning Two New Sciences,* Dover Publications, New York, 1952.

D. A. Greenberg, *Mathematics for Introductory Science Courses: Calculus and Vectors,* W. A. Benjamin, New York, 1965.

E. R. Huggins, *Physics I,* W. A. Benjamin, New York, 1968. Chapters 3 and 4 present a clear discussion of vectors.

M. Jammer, *Concepts of Space,* Harper and Row, New York, 1960.

G. Kayafas, *Stopping Time: The Photographs of Harold Edgerton,* published by Harry N. Abrams, Inc. (paperback), 2000, ISBN: 0810927179. A bullet seen the instant it explodes through an apple...a milk-drop splash transformed into a perfect coronet...a golf swing shown in a procession of exposures taken milliseconds apart—some of the world's most famous photographs were taken by Harold Edgerton, the late MIT scientist who invented the electronic flash. This great classic, first published in 1987 and now back in print in paperback, celebrates Edgerton's genius and his lasting influence on photography.

J. Piaget, *The Child's Conception of Geometry,* Basic Books, New York, 1960.

J. Piaget, *The Child's Conception of Space,* Humanities Press, New York, 1948.

The interaction concept is being used more and more widely to explain social and scientific phenomena. At conferences, strong interaction may be evident among some participants, weak interaction among others. At the ocean shore, erosion is caused by the interaction of wind and water with rock. In the laboratory, magnets interact even when they are not touching.

A dictionary provides the following definitions:

interact (verb): to act upon each other ...;

interaction (noun): action upon or influence on each other.

To say that objects interact, therefore, is to say that they have a relationship wherein they jointly produce an effect, which is the result of their action upon each other. In the examples cited above, anger may be the effect caused by strong (and irritating) interaction among the conference participants; crumbling and wearing away is the effect of the interaction of wind and water with rocks; and movement toward one another followed by sticking together is the effect of the interaction of the magnets.

3.1 Evidence of interaction

We take the point of view that influence and interaction are abstractions that we cannot observe directly. What we *can* observe are the effects or results of interaction. Congressional passage of unpopular legislation requested by the President would be an observable effect of the President's influence and therefore would be called evidence of his influence. The change in direction of motion of a struck baseball is an observable effect of its interaction with the bat and therefore can be called evidence of interaction.

You may believe that you can sometimes observe the interaction itself, as when a bat hits a baseball or a typewriter prints a letter on a piece of paper. These examples, which include physical contact and easily recognized effects, seem different from those where magnets interact without contact or where erosion is so slow that the effects are imperceptible. This apparent difference, however, is an illusion. You observe only the close proximity of bat and ball, a sound, and the change of the ball's direction of motion, all of which are so closely correlated and so familiar that you instantly interpret them as evidence of interaction between bat and ball. The interaction of bat and ball is, however, merely the relationship whereby the observable effects are brought about, and relationships are abstractions that cannot be observed directly.

Indirect evidence of interaction. An example where the evidence is very indirect was described in Section 2.2. In his analysis of the films of President Kennedy's assassination, Alvarez interpreted a blurred photograph as evidence of interaction between the photographer and a rifle being fired. No one can question the blurring in the photograph, which is directly observable. But it is clear that not everyone may agree with Alvarez's interpretation that there is a relationship between the blur and a rifle shot.

Another example of interpretation of indirect evidence for interaction is in the relationship between cigarette smoking and lung cancer. Lung cancer and cigarette smoking are separately observable, and the statistical evidence from the 1950s showed a very strong correlation, sufficient to warrant the conclusion of a pathological interaction between cigarette smoke and lung tissue (publicized in the Surgeon General's Report for 1964). There is now, in 2003, a much larger body of evidence for this interaction. Yet many smokers do not take this interpretation of the evidence seriously enough to believe that they are slowly committing suicide.

Alternate interpretations. The critical problem in interpreting evidence of interaction is that any one of several different interactions might be responsible for the same observed effect. The typed letter in the example of the typewriter and the paper does not furnish conclusive evidence as to which typewriter made the letter, a question that sometimes arises in detective stories. A direct way to overcome this weakness is to find evidence that supports one hypothesis. For instance, the paper might be beside a typewriter, the ribbon on one typewriter might match the shade of the typed letter, or a defect in the machine's type might match one that appears in the typed letter. If so, the original identification of the typewriter is supported.

An indirect way to support one hypothesis is to eliminate alternatives. By checking many typewriters and finding how poorly they match the ribbon color and type impression, the detective may be able to eliminate them from further consideration. Supporting evidence for one alternative and/or evidence against other alternatives will enable you to establish one hypothesis conclusively or may only lead you to decide that one of them is more likely than the others. A procedure for finding such evidence by means of control experiments is described in Section 3.4.

When you suspect that there might be interaction, you should make a comparison between what you observe and what you would expect to observe in the absence of interaction. If there is a difference, you can interpret your observation as evidence of interaction and seek to identify the interacting objects; if there is no difference, you conclude that there was no interaction or that you have not observed carefully enough.

3.2 Historic background

Mankind has not always interpreted observed changes or discrepancies as evidence of interaction. In ancient times, some philosophers took the view that changes were brought about by a fate or destiny that was inherent in every object. In our own day, many people ascribe specific events to supernatural or occult forces. These forms of explanation and the interaction concept we are advocating, however, *do* share a common feature: they are both attempts to explain regular patterns in nature so as to anticipate the future and possibly to influence and control future events.

Cause and effect. When you observe two happenings closely correlated in space and time, you tend to associate them as cause and effect. Such a conclusion is reinforced if the correlation of the happenings persists in a regular pattern. The person who strikes a match and observes it bursting into flame infers that the striking caused the fire. The primitive man who performs a rain dance infers that the dance causes the ensuing rain. Even the laboratory pigeon that receives a pellet of grain when it pecks a yellow card becomes conditioned to peck that card when hungry. These individuals will repeat their actions - striking the match, dancing, or pecking the card - if they wish to bring about the same consequences again. After a sufficient number of successful experiences, all three will persist in their established behavior, even though some failures accompany their future efforts.

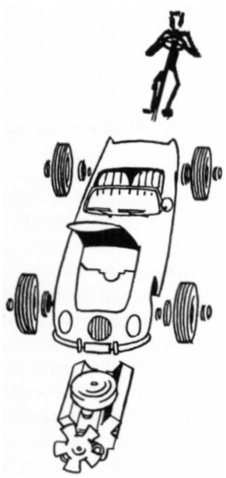
The interaction viewpoint. We may state the distinction between the modern scientific approach and other types of explanations for events in the following way: the scientist ascribes happenings to interactions among two or more objects rather than to something internal to any one object. Thus, the falling of an apple is ascribed to its gravitational interaction with the earth and not to the heaviness inherent in the apple. The slowing down of a block sliding on a table is ascribed to friction between the block and the table and not to the power or "desire" of the block to come to rest by itself. Fire is the manifestation of combustion, that is, the interaction of fuel and oxygen, and is not itself an element. The rain dance and the rain, however, cannot be put into this framework; therefore, this association is nowadays considered a superstition.

At any one time, however, science cannot provide explanations for all possible happenings. When a new phenomenon is discovered, the interacting objects responsible for it must be identified, and this may be difficult. The origin of some of the recently discovered radiation reaching the earth, for example, is yet to be found. On the other hand, we now have identified specific chemical substances in cigarette smoke that cause cancer. We are beginning to understand the specific biochemical mechanisms by which such substances cause lung cells to start the explosive multiplication that manifests itself as cancer. Research is continuing to further elucidate the details of this dangerous interaction.

3.3 Systems

The word "system" has entered our daily lives. Communication systems, computer systems, and systems analysis are discussed in newspapers and magazines and on television. In all these discussions, and in this text as well, the word "system" refers to a whole made of parts.

The systems concept is applied whenever a whole, its parts, and their inter-relationships must all be kept clearly in mind, as illustrated in the following two examples. Traffic safety studies take into account an entire driver-car system and do not confine themselves merely to the engineering of the car or the health of the driver. A physician realizes that the human heart, though a single organ, is really a complex system composed of muscles, chambers, valves, blood vessels, and so on. The system is physically or mentally separated from everything else so that the relations among the parts may be studied closely.



To simplify our terminology, we will often refer to the whole as "system" and to the parts as "objects." Thus, the car and the driver are the objects in the driver-car system, and the muscles, chambers, and so on are objects in the system called "the heart." By using the word "object" to refer to any piece of matter (animate or inanimate, solid, liquid, or gaseous), we are giving it a broader meaning than it has in everyday usage.

Sometimes one of the parts of a system is itself a system made of parts, such as the car (in the driver-car system), which has an engine, body, wheels, and so forth. In this case, we should call the part a subsystem, which is a system entirely included in another system.

In a way, everyone uses the systems concept informally, without giving it a name. At times, everyone focuses attention temporarily on parts of the environment and ignores or neglects other parts because the totality of incoming impressions at any one moment is too complex and confusing to be grasped at once. The system may have a common name, such as "atmosphere" or "solar system," or it may not, as in the example of the jet fuel and liquid oxygen that propel a rocket. The systems concept is particularly useful when the system does not have a common name, because then the group of objects under consideration acquires an individual identity and can be referred to as "the system including car and driver" or more briefly as "the driver-car system" once the parts have been designated.

"... in all the operations of art and nature, nothing is created; an equal quantity of matter exists both before and after the experiment, ... and nothing takes place beyond changes and modifications in the combinations of these elements. Upon this principle, the whole art of performing chemical experiments depends."

Antoine Lavoisier
Traite Elementaire de Chimie, *1789*

Conservation of systems. Once we have identified a system, changes may occur in the system. We must have a way to identify the system at later times in spite of the changes. A chemist uses the conservation of matter to identify systems over time. This means that no matter can be added to or removed from the matter originally included in the system. For example, when jet fuel burns, the fuel and oxygen become carbon dioxide and water. Therefore, the chemist thinks of the carbon dioxide and water as being the same system as the jet fuel and oxygen, even though the chemical composition and temperature have changed.

The psychotherapist and the economist do not use the same criteria as the chemist for following the identity of a system over time. The psychotherapist focuses his attention on a particular individual with a personality, intellectual aptitudes, and emotions. A therapist, therefore, selects this individual as a system that is influenced by its interaction with other individuals and by its internal development. The person as a system retains its identity even though it exchanges matter with its environment (breathing, food consumption, waste elimination). For the economist, all the production, marketing, and consuming units in a certain region constitute an economic system that retains its identity even though persons may immigrate or emigrate and new materials and products may be shipped in or out.

The physicist studying macro-domain phenomena finds the matter-conserving system most useful. This is, therefore, the sense in which we will use the systems concept throughout this text. In the micro domain, however, the concepts of matter and energy have acquired new meanings during the last few decades, and, if you study physics further, you will learn how to expand and modify these criteria for defining systems.

Place two identical pieces of clean writing paper in front of you. Pick up one piece and call it System P.

(1) Wrinkle the paper in your hand into a ball. Is what you now hold in your hand System P?

(2) Is the paper lying on the table System P?

(3) Tear the wrinkled paper in half, and hold both pieces. Is what you now hold in your hand System P?

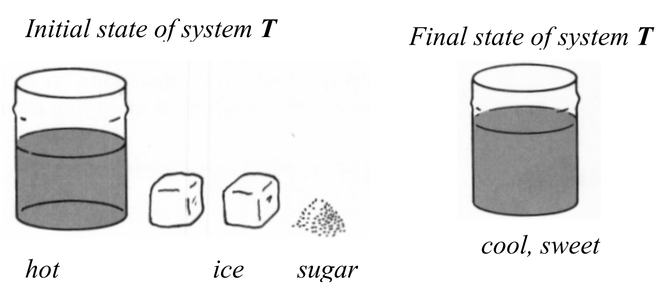
(4) Put down one of the two torn pieces. Is what you now hold in your hand System P?

You can apply conservation of matter to the selection of systems in two ways. By watching closely, you can determine whether you see the same system before and after an event. For instance, when a bottle of ginger ale is opened, some of the carbon dioxide gas escapes rapidly. The contents of the sealed bottle (we may call it System A), therefore, are not the same system as the contents of the opened bottle, which may be called System B. The escaping bubbles are evidence of the loss of material from the bottle.

In the second kind of application, you seek to keep track of the system even though its parts move from one location to another. Thus, after the bottle is opened, System A consists of System B plus the escaped gas; the latter, however, is now mixed with the room air and can be conveniently separated from the room air only in your mind. For this reason we stated at the beginning of this section that a system of objects need only be separated mentally from everything else; sometimes the physical separation is difficult or impossible to achieve, but that is immaterial for purposes of considering a system.

State of a system. To encompass the continuity of the matter in the system as well as the changes in form, it is valuable to distinguish the identity of the system from the state of the system. The identity refers to the material ingredients, while the state refers to the form or condition of all the material ingredients (Fig. 3.l). . Variable factors, such as the distance between objects in the system, its volume, its temperature, and the speeds of moving objects, are used to describe the state. In

Figure 3.1 Change in the state of a system.



Chapter 4 we will relate matter and energy, which are of central concern to the physical scientist, to changes in the state of a system. There we will describe the ways in which a system may store energy and how energy may be transferred as changes occur in the state of a system. From an understanding of energy storage and transfer has come the extensive utilization of energy that is at the base of modern technology and current civilization.

Investigations of interacting objects. In their research work, physicists study systems of interacting objects in order to classify or measure as many properties of the interactions as they can. They try to determine which objects are capable of interacting in certain ways, and which are not (e.g., magnetic versus nonmagnetic materials). They try to determine the conditions under which interaction is possible (a very hot wire emits visible light but a cold wire does not). They try to determine the strength of interaction and how it is related to the condition and spatial arrangement of the objects (a spaceship close to the earth interacts more strongly with the earth than does one that is far away from the earth). Physicists try to explain all physical phenomena in terms of systems of interacting objects or interacting subsystems.

Working models for systems and the structure of matter. There are some happenings, however, such as the contraction of a stretched rubber band that involve only a single object and appear to have no external causes. In such cases, the scientist makes a working model in which the object is made of discrete parts. A working model for the rubber band is made of parts called "rubber molecules." The properties of the entire system are then ascribed to the motion and the interaction of the parts. Some models are very successful in accounting for the observed behavior of the system and even suggest new possibilities that had not been known but that are eventually confirmed. Such a model may become generally accepted as reality: for instance, everyone now agrees that rubber bands are systems made of rubber molecules. Also, further model building may represent the rubber molecules as subsystems composed of parts called "atoms" and explain the behavior of the molecules in terms of the motion and interaction of the atoms.

This kind of model building is called the search for the structure of matter - how ordinary matter in the macro domain is composed of interacting parts, and these parts in turn are composed of interacting parts, and so on into the micro domain. One of the frontiers of science is the search for ultimate constituents, if such exist. Since we will always find more questions to ask, it is unlikely that we will ever accept the concept of an "ultimate constituent."

3.4 Collecting evidence of interaction

Interactions are recognized by their effects, that is, by the difference between what is actually observed and what would have been observed in the absence of interaction. Such a difference is evidence of interaction. The systems concept is of great value here because it enables you to designate and set apart (at least mentally) the objects that are being compared as you look for a difference. One approach is to compare a system before an event (in its so-called initial state) with the same system after an event (in its so-called final state). For example, you compare a section of bare skin on the morning and the evening of a day at the beach (Fig. 3.2). The section of skin is the system. In this experiment you assume based on your experience that the skin color would not have changed in the absence of interaction. The observed change in skin color is therefore evidence of interaction with the sun.

As another example, take some sugar and let it dissolve in water in a glass beaker to form a solution (Fig 3.1). At the beginning of the experiment, the water-sugar system consists of dry crystals and colorless, tasteless water. At the end, there are no crystals and the liquid tastes sweet. The change in the state of this system is evidence of interaction between sugar and water.

Control experiment. Consider now an experiment in which you put yeast into a sugar solution in a glass and let this system stand in a warm place for several days. You will observe bubbles, an odor, and a new taste - that of ethyl alcohol. These changes can be interpreted as evidence of interaction within the water-sugar-yeast system. Can you narrow down the interacting objects more precisely or are all three parts necessary?

For comparison, suppose you can conduct experiments in which one ingredient is omitted. You dissolve sugar in water without yeast, you dissolve yeast in water without sugar, and you mix sugar and yeast. Each of these is called a control experiment; from their outcomes, you can answer the question above. By designing other control

Figure 3.2 The skin shows evidence of interaction with the sun only where it was exposed to sunlight. The exposed skin can be compared to the unexposed areas.

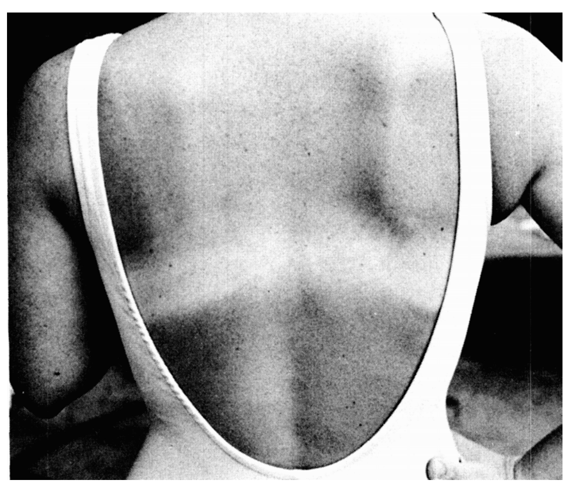


Figure 3.3 When you try to determine which electric circuit breaker supplies power to a particular light fixture, you turn on the switch of the fixture and then turn off the circuit breakers one at a time. In one of these "experiments" the bulb darkens, in the others it does not. Each turning off serves as a control experiment to be compared to the situation in which all circuits are turned on.

"It frequently happens, that in the ordinary affairs and occupations of life, opportunities present themselves of contemplating some of the most curious operations of Nature . . . 1 have frequently had occasion to make this observation; and am persuaded, that a habit of keeping the eyes open to everything that is going on in the ordinary . . . business of life has oftener led, as it were by accident . . . to useful doubts and sensible schemes for investigation and improvement, than all the most intense meditations of philosophers in the houses expressly set apart for study."
Benjamin Thompson,
Count Rumford
Philosophical Transactions, 1798

experiments, you can try to determine whether the glass container was necessary, and whether the temperature of the environment made any difference.

By carrying out control experiments, you try to identify those objects in the system that interact and those whose presence is only incidental (Fig. 3.3).

Inertia. One other important concept in the gathering of evidence of interaction is the concept of inertia. Inertia is the property of objects or systems to continue as they are in the absence of interaction, and to show a gradually increasing change with the elapse of time in the presence of interaction. For example, you expected the pale skin on the girl's back (Fig. 3.2) to remain pale as long as it was not exposed to the sun. You expect a rocket to remain on the launching pad unless it is fired. You expect sugar crystals to retain their appearance if they are not heated, brought into contact with water, or subjected to other interactions. You expect an ice cube to take some time to melt even when it is put into a hot oven.

Your everyday experience has taught you a great deal about inertia of the objects and systems in your environment. When you compare the final state with the initial state of a system and interpret a difference as evidence of interaction, you are really using your commonsense background regarding the inertia of the system. You must be careful, however, because occasionally your commonsense background can be misleading.

"... we may remark that any velocity once imparted to a moving body will be rigidly maintained as long as the external causes of acceleration or retardation are removed . . ."
Galileo Galilei
Dialogues Concerning Two New Sciences, 1638

Inertia of motion. The motion of bodies also exhibits inertia. Curiously, motion is one of the most difficult subjects to treat scientifically because of commonsense experience. When you see a block gliding slowly on an air track (Fig. 3.4), you almost think it must contain a motor because you expect such slowly moving objects to come to rest after a very short time. In fact, the block is only exhibiting its inertia of motion because the frictional interaction with the supporting surface is very small. You must, therefore, extend your concept of inertia to cover objects in motion (such as the block), which tend to remain in motion and only gradually slow down if subject to a frictional interaction. You must also extend it to objects at rest (such as the rocket), which tend to remain at rest and only gradually acquire speed if subject to an interaction. Change in speed from one value to another - where the state of rest is considered to have "zero" speed - is therefore evidence of interaction. Galileo already identified inertia of motion even though he did not give it a name. Isaac Newton framed a theory for moving bodies in which he related their changes in speed and direction of motion to their interactions. The "laws of motion," as Newton's theory is called, will be described in Chapter 14.

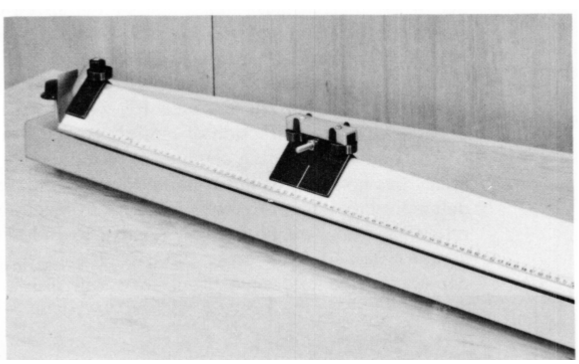
OPERATIONAL DEFINITION
Inertial mass is measured by the number of standard units of mass required to give the same rate of oscillation of the inertial balance.

A key concept in the laws of motion is that of the *inertial mass*. This is an extension of Galileo's idea that, in the absence of external influences, objects maintain their state of motion, whether at rest or moving; it is useful to have a numerical quantity which measures the extent to which an object does this: "inertial mass" is the name for this quantity. Speaking roughly, inertial mass is the degree to which a body tends to maintain its state of motion. More specifically, an object with a large inertial mass takes longer to speed up (or slow down) than an object with small inertial mass. It is important to keep in mind the *difference* between *inertial* mass and *gravitational* mass. The latter (Section 1.5) is connected with the downward pull of gravity (the weight) and can be measured with an equal-arm balance. In contrast, inertial mass can be defined and measured with a device called the *inertial balance* (Fig. 3.5) to compare two objects or to compare an object of unknown inertial mass with standard units of inertial mass.

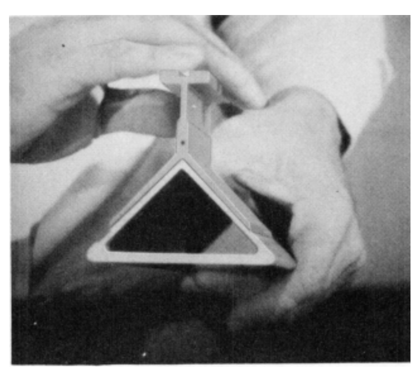
The inertial balance operates *horizontally*, thus eliminating the effects of gravity. The body attached to the end of the steel strip is repeatedly speeded up and slowed down by the oscillation of the strip. The inertia of the body, therefore, strongly influences the rate

Figure 3.4 An air track (below).
(a, below left) Small holes in the track emit tiny jets of air. When a close fitting metal piece passes over an opening, the air is trapped and forms a thin film over which the metal piece can slide with very little friction.
(b, below right) The closeness of fit can be seen in this end view.

(a)



(b)



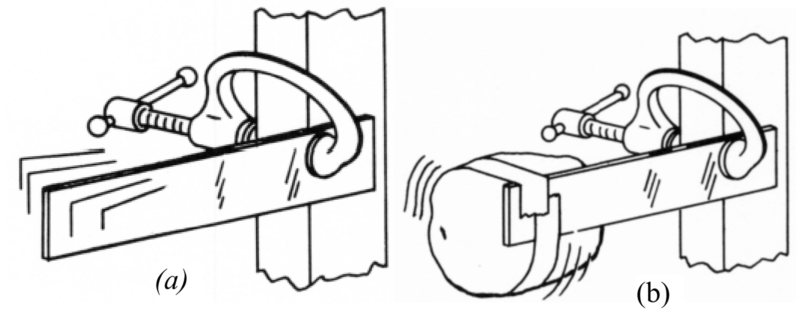


Figure 3.5 The inertial balance.
(a) The inertial balance consists of an elastic steel strip, which oscillates back and forth after the free end is pulled to the side and released.
(b) When objects are attached to the end of the strip, the oscillations take place more slowly. The inertial mass of a stone is equal to the number of standard objects required to give the same count of oscillations per minute. To measure the inertial mass of the stone, it is attached to the end of the steel strip and set into oscillation. The number of oscillations in 1 minute is counted. Then the stone is taken off, a number of standard objects are attached, and their number adjusted until the count of oscillations is equal to the count obtained with the stone.

of oscillation of the strip: large inertia (resistance to change of speed) means slow oscillations, small inertia means rapid oscillations.

The generally accepted standard unit of inertial mass is the kilogram, represented by the same platinum-iridium cylinder as the unit of gravitational mass. Even though inertial and gravitational masses are measured in the same units, they are different concepts and have different operational definitions. Inertial and gravitational mass are both important for understanding motion, particularly bodies falling under the influence of gravity. We will focus on this in Chapter 14.

FORMAL DEFINITION
Momentum is the product of inertial mass multiplied by instantaneous speed.

The unit of momentum does not have a special name; it is a composite unit, kilogram-meters per second (kg m/sec) that combines mass and speed.

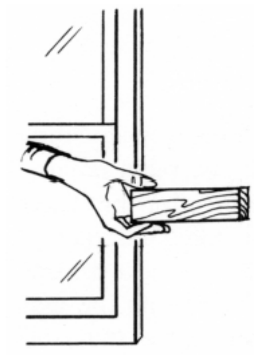
Equation 3.1
momentum $= \mathcal{M}$
speed $= v$
inertial mass $= M_I$

$$\mathcal{M} = M_I \ v$$

A second important concept in the laws of motion is the *momentum* of a moving body. The word is commonly applied to a moving object that is difficult to stop. A heavy trailer truck rolling down a long hill may, for instance, acquire so much momentum that it cannot be brought to a stop at an intersection at the bottom. By contrast, a bicycle coasting down the same hill at the same speed has much less momentum because it is less massive than the truck.

The physical concept of *momentum* is defined formally as the product of the inertial mass multiplied by the speed of the moving object (Eq. 3.1). This concept was used by Newton to formulate the laws of motion (Chapter 14), and it plays an important role in the modern models for atoms (Sections 8.3, 8.4, and 8.5). We will elaborate on the momentum concept in Chapter 13, where we will describe how it depends upon the direction of motion as well as on the speed.

If we want to use changes of motion as evidence of interaction, we must be careful because, as we have pointed out in Chapter 2, motion must be defined relative to a reference frame. An object moving relative to one reference frame may be at rest relative to another. Evidence of interaction obtained from observation of moving objects, therefore, will depend on the reference frame. We will ordinarily use a reference frame attached to a massive body such as the earth (for terrestrial phenomena) or the sun (for the solar system).



Combined interaction. A block held in your hand does not show evidence of interaction (i.e., it remains at rest), yet it is clearly subject to interaction with the hand and with the earth. This is an example of what we must describe as two interactions combining in such a way that they compensate for one another and give the net effect of no interaction. Situations such as this raise the question of the strength of interaction; how can you compare two interactions to determine whether they can compensate exactly or not, other than to observe their combined effect on the body? We will take up this question in Chapter 11.

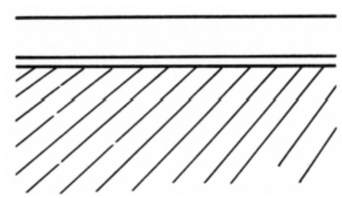
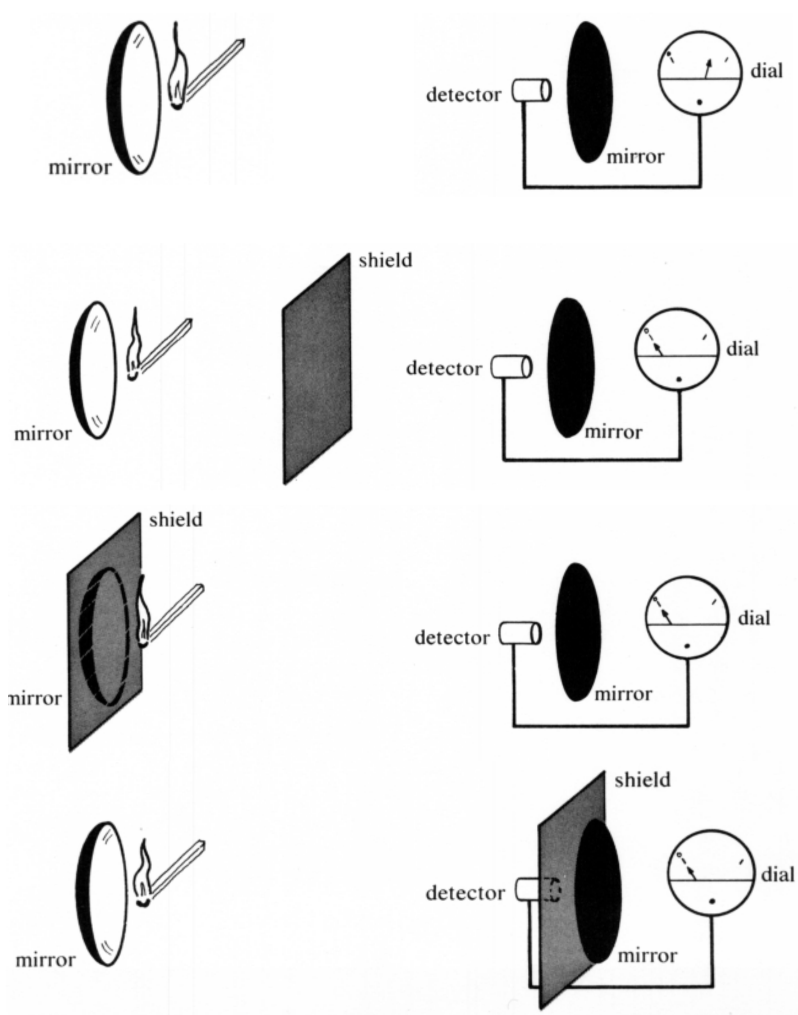
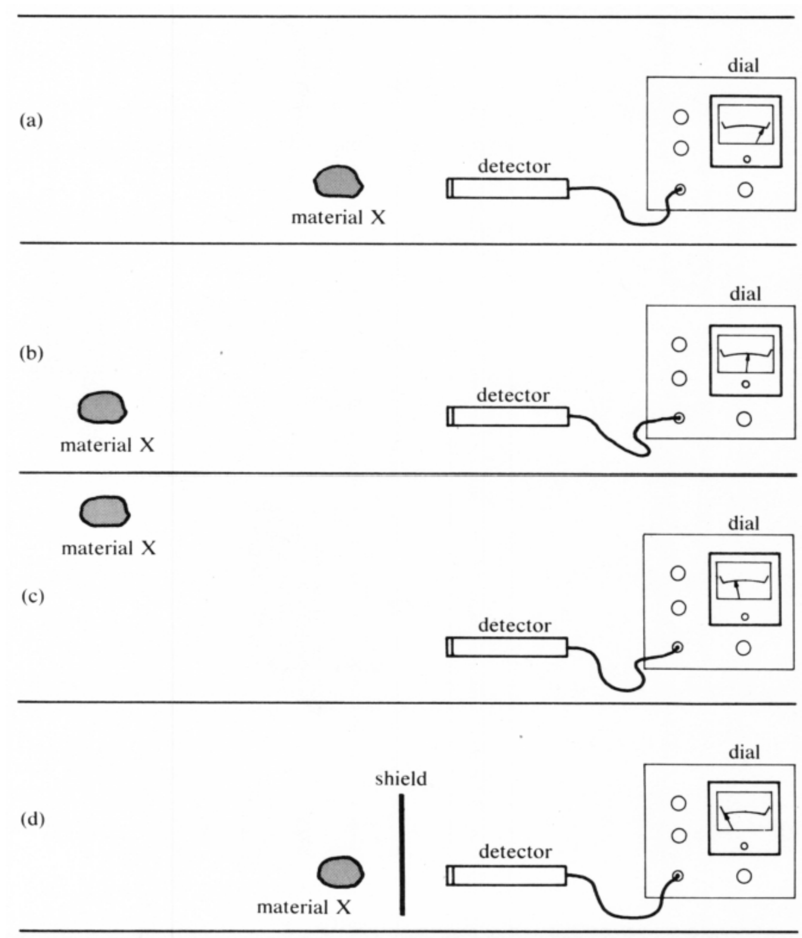


Figure 3.6 Four steps in the investigation of the interaction of a match flame with a detector show the effects of a shield placed in various locations.



Radiation. A situation that contains a different element of mystery is illustrated in Fig. 3.6. Two mirrors are facing each other at a separation of several meters. There is no mechanical connection between them. At a central point near one mirror is a device called a detector, which is connected to a dial. If a lighted match is placed at a central point in front of the opposite mirror, you see a deflection on the dial (Fig. 3.6a). After a little experimentation, you recognize that the placement of the match and the dial deflection are definitely correlated. This is the evidence of interaction between the match and the detector. If, now, the match is held in position and a cardboard shield is placed in various positions in the apparatus, the deflection falls to zero [Figs. 3.6 (b), (c), and (d)]. Without anyone touching any of the visible objects used for the experiment, an effect was produced. The inference is that something

Figure 3.7 Four steps in an investigation of material X show evidence of interaction between the material and the detector.



was passing from the match to the detector by way of the mirrors, and that the shield somehow blocked or interrupted this passage.

We, therefore, construct a working model that is just like the experimental system but includes in addition an "object" that passes from the match to the first mirror, the second mirror, and the detector. The scientist calls this "model object" radiation. In terms of this model, he can describe the effect of the shield on the dial reading as evidence of interaction between the shield and the radiation, he can describe the path of the radiation, he can describe the match as a radiation source, and he can describe the detector as a radiation detector.

Another experiment, with a rocklike material X and a detector with a dial, is illustrated in Fig. 3.7. From the evidence you may conclude that material X is not an ordinary inert rock but is a source of radiation, and you make a working model that includes an "object," again called radiation, that passes between material X and the detector. After this discovery, you can study the spatial distribution of the radiation by holding the detector in various directions and at various distances from the rock, you can study the interaction of the radiation with various shields (cardboard, glass, iron, and aluminum) placed to intercept it, and so on. From this kind of investigation you become more familiar with the radiation from material X and may, eventually, think of it as a real object and not only as part of a model.

The discovery of evidence of interaction is a challenge to identify the interacting objects and to learn more about the interaction: the conditions under which it occurs, the kind of objects that participate, the strength and speed with which the evidence appears, and so on. It can be the beginning of a scientific investigation.

3.5 Interaction-at-a-distance

Consider now a common feature of the two experiments with radiation. In both cases, you observed evidence of interaction between objects that were not in physical contact. We speak of this condition as *interaction-at-a-distance* because of the distance separating the interacting objects. The idea that objects interact without touching seems to contradict our intuition based on physical experience and the sensations of our bodies; therefore, we construct working models that include radiation to make interaction possible between the two objects. The shields intercept the radiation and show the effect of its presence and absence; this confirms the usefulness of our working models.

An experiment that significantly resembles the radiation experiments can be carried out with the system shown in Fig. 3.8. A spring is supported at the ends by rigid rods. If a ruler strikes the spring at point A, you see a disturbance in the spring, which is evidence of interaction, and then movement of the flag at B, another piece of evidence of interaction. The first movement is evidence of interaction between the ruler and the spring. The second is evidence of interaction of the spring with the flag.

The experiment with the spring and the flag becomes another example of interaction-at-a-distance, however, if you choose to focus

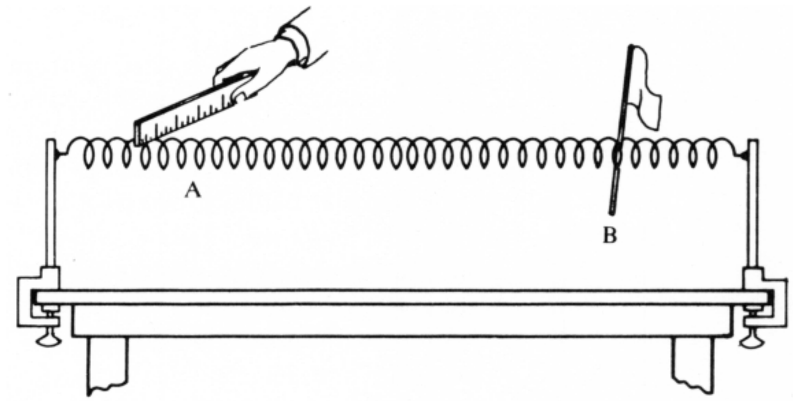


Figure 3.8 The ruler interacts with the flag by way of the long spring. Is this an example of interaction-at-a-distance?

on the system including only the ruler and the small flag. The motion of the flag correlated with the motion of the ruler is evidence of interaction-at-a-distance between these two objects. Of course, in this experiment you can see the spring and a disturbance traveling from the ruler along the spring to the flag. You do not need to construct a working model with a "model object" to make the interaction possible. You can, therefore, use the disturbance along the spring as an analogue model to help you visualize the radiation traveling from the match or material X to the detectors in the two other experiments.

The field model. Familiar examples of interaction-at-a-distance are furnished by a block falling toward the earth when it is not supported, by a compass needle that orients itself toward a nearby magnet, and by hair that, after brushing on a dry day, extends toward the brush. The intermediaries of interaction-at-a-distance in all these examples are called *fields*, with special names, such as *gravitational field* for the block-earth interaction, *magnetic field* for the compass needle-magnet interaction, and *electric field* for the brush-hair interaction. We may call this approach the *field model* for interaction-at-a-distance.

Radiation and fields. Do radiation and fields really exist, or are they merely "theoretical objects" in a working model? As we explained in Section 1.3, the answer to this question depends on how familiar you are with radiation and fields. Since radiation carries energy from a source to a detector, while the field does not accomplish anything so concrete, radiation may seem more real to you than fields. Sunlight, the radiation from the sun to green plants or to the unwary bather, is so well known and accepted that it has had a name for much longer than has interaction-at-a-distance. Nevertheless, as you become more familiar with the gravitational, magnetic, and electric fields, they also may become more real to you.

"The physicist ... accumulates experiences and fits and strings them together by artificial experiments ... but we must meet the bold claim that this is nature with ... a good-humored smile and some measure of doubt."
Goethe
Contemplations of Nature

For the scientist, both radiation and fields are quite real. In fact, the two have become closely related through the field theory of radiation, in which the fields we have mentioned are used to explain the production, propagation, and absorption of radiation. More on this subject is included in Chapter 7.

Gravitational field. Two fields, the gravitational and the magnetic, are particularly familiar parts of our environment. At the surface of the earth the gravitational field is responsible for the falling of objects and for our own sense of up and down. The plumb line (Section 1.4) and the equal-arm balance (Section 1.5) function because of the gravitational interaction between the plumb bob or the weights and the earth. We, therefore, use a plumb line to define the direction of the gravitational field at any location. Because the earth is a sphere, the direction of the gravitational field varies from place to place as seen by an observer at some distance from the earth (Fig. 3.9). More about the gravitational field will be described in Chapter 11.

OPERATIONAL DEFINITION
The direction of the gravitational field is the direction of a plumb line hanging freely and at rest.

Magnetic field. The magnetic field is explored conveniently with the aid of a magnetic compass, which consists of a small, magnetized needle or pointer that is free to rotate on a pivot (Fig. 3.10). When the compass is placed near a magnet, the needle swings back and forth,

Figure 3.9 (to right) The gravitational field near the earth is directed as indicated by plumb lines. The field appears to converge on the center of the earth

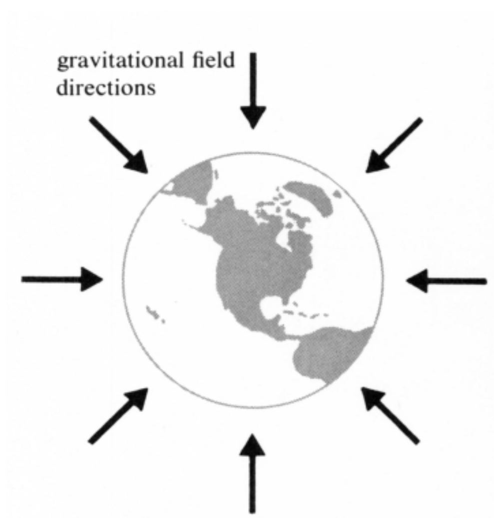
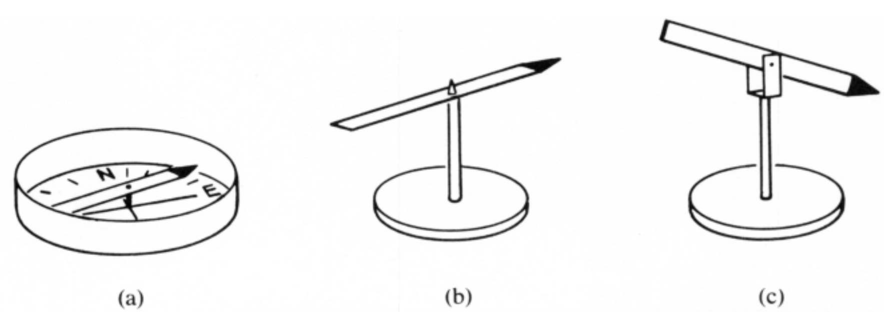


Figure 3.10 Examples of magnetic compasses. (below)
(a) The compass needle is often enclosed in a case for better protection.
(b) The pivot may permit the needle to rotate in a horizontal plane.
(c) The pivot may permit the needle to rotate in a vertical plane.



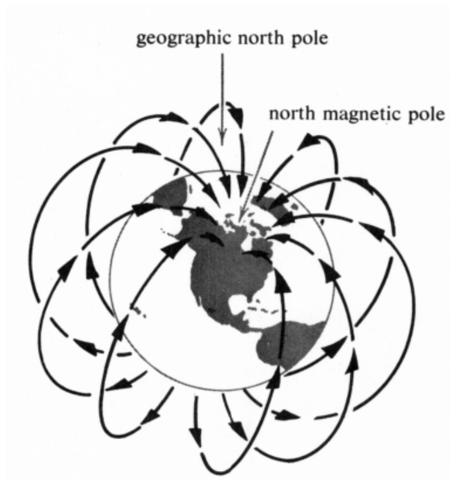


Figure 3.11 The magnetic field of the earth, represented by the arrows, lies close to the geographic north-south direction, but does not coincide with it. In the magnetic dipole model for the earth, the two magnetic poles lie near the center of the earth on a line through northern Canada and the part of Antarctica nearest Australia.

and finally comes to rest in a certain direction. Because of its interaction with other magnets, the compass needle functions as detector of a magnetic field at the point in space where the compass is located. It is most commonly used to identify the direction of the magnetic field at the surface of the earth, which lies close to the geographic north-south direction (Fig. 3.11). Since the compass needle has two ends, we must decide which end indicates the direction of the magnetic field. The accepted direction of the magnetic field is that of the geographic north-seeking end of the needle (henceforth called the "direction of the needle"), as shown by the arrows in Fig. 3.11**.**

OPERATIONAL DEFINITION
The direction of the magnetic field is the direction of a compass needle that is free to rotate and has come to rest.

Figure 3.12 The arrows represent the compass needles that indicate the magnetic field near the bar magnet.

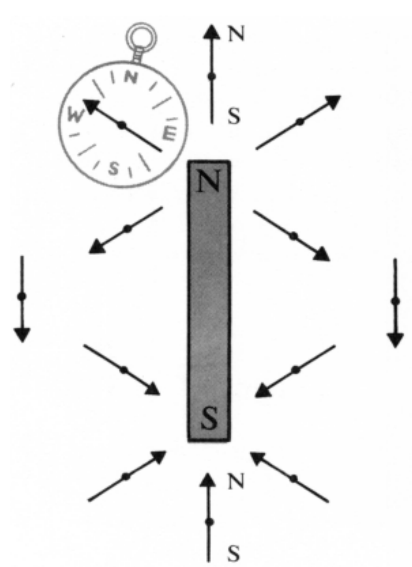
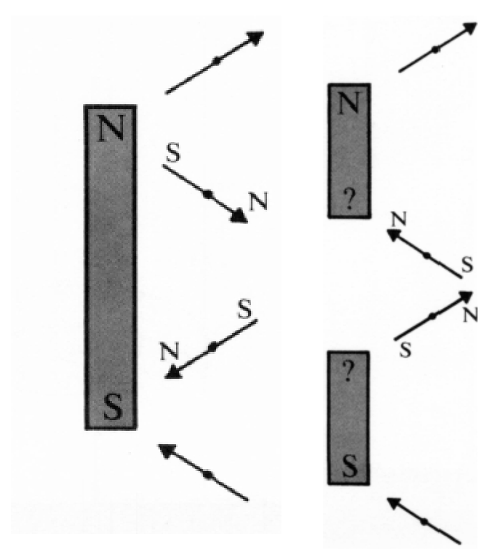


Figure 3.13 A bar magnet is cut in half in an effort to separate the north pole from the south pole. Arrows represent compass needles. Each broken part still exhibits two poles, the original pole and a new one of the opposite kind.



William Gilbert (1544-1603), an Elizabethan physician and scientist, wrote the first modern treatise on magnetism, De Magnete*. Gilbert worked with natural magnets (lodestones). In one chapter of this work, Gilbert introduced the term electric (from the Greek elektron for amber).*

Strength of the magnetic field. When you place a compass near a magnet, you notice that the needle swings back and forth rather slowly if it is far from the magnet and quite rapidly if it is close to the magnet. You can use this observation as a rough measure of interaction strength or magnetic field strength: rapid oscillations are associated with a strong field, slow oscillations with a weak field. You thereby discover that the magnetic field surrounding a magnet has a strength that differs from point to point; the field strength at any one point depends on the position of that point relative to the magnet.

Magnetic pole model. When you explore the magnetic field near a bar magnet, you find that there are two regions or places near the ends of the magnet where the magnetic field appears to originate. This common observation has led to the magnetic pole model for magnets as described in 1600 by William Gilbert (quoted to left). In this model, a magnetic pole is a region where the magnetic field appears to originate. The magnetic field is directed away from north poles and toward south poles according to the accepted convention (Fig. 3.12). All magnets have at least one north pole and one south pole. Opposite poles of two magnets attract one another, and like poles repel. If you apply these findings to the compass needle itself, you conclude that the north-seeking end of the needle contains a north pole (it is attracted to a magnetic south pole, Fig. 3.12).

"...thus do we find two natural poles of excelling importance even in our terrestrial globe . . . In like manner the lodestone has from nature its two poles, a northern and a southern . . . whether its shape is due to design or to chance . . . whether it be rough, broken-off, or unpolished: the lodestone ever has and ever shows its poles."
William Gilbert
De Magnete, 1600

An obvious question now suggests itself: can a magnetic pole be isolated? So far, physicists have failed in all their attempts to isolate magnetic poles (Fig. 3.13), in that they have not been able to narrow down the regions inside magnets where the magnetic field originates. They have found instead that the magnetic field appears to continue along lines that have no beginning or end but loop back upon themselves (Fig. 3.14). Thus, magnetic poles appear to be useful in a working model for magnets when the magnetic field outside magnets is described, but they fail to account for the field inside magnets.

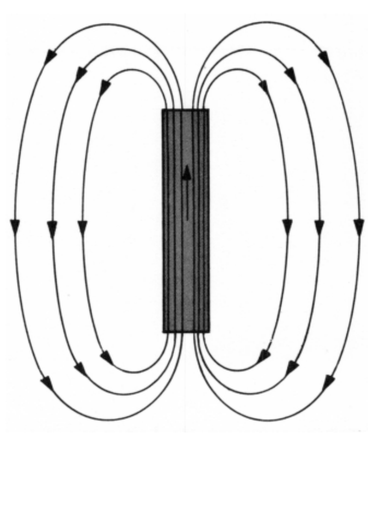


Figure 3.14 So-called magnetic field lines indicate the direction of the magnetic field. Lines inside the magnet close the loop made by the lines outside the magnet.

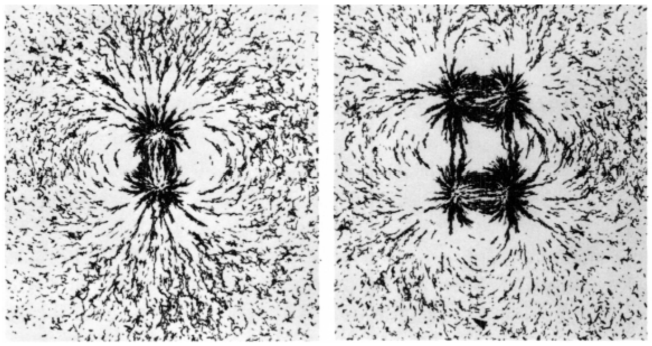


Figure 3.15 Iron filings were sprinkled over a piece of paper that concealed one or two bar magnets.

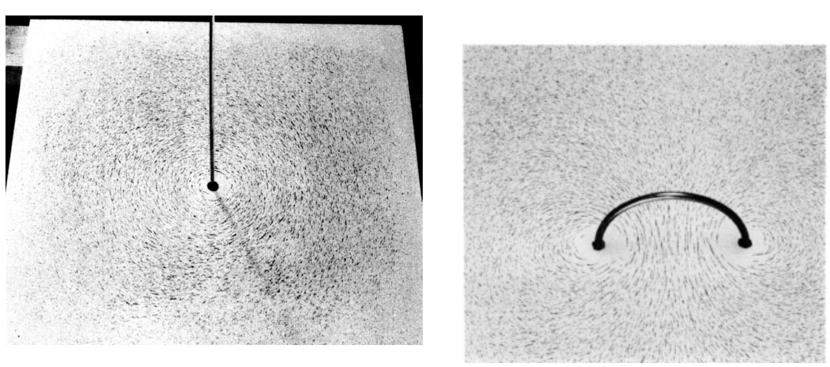
Hans Christian Oersted (1777-1851) inherited his experimental acumen from his father, an apothecary. His famous discovery in 1819 of the magnetic field accompanying an electric current took place as he was preparing for a lecture demonstration for his students. Until that time, electricity and magnetism were considered unrelated. Thus Oersted's discovery initiated intensive study of the relationships between electricity and magnetism, and these two separate disciplines gradually merged into the branch of physics known as electromagnetism.

Display of the magnetic field. Another technique for exploring a magnetic field is to sprinkle iron filings in it (Fig. 3.15). The filings become small magnets and, like the compass needle, tend to arrange themselves along the direction of the magnetic field. They produce a more visual picture of the magnetic field. This method is less sensitive than the compass, because the filings are not so free to pivot.

Electromagnetism. Not quite 150 years ago, Hans Christian Oersted, while preparing for a lecture to his students, accidentally found evidence of interaction between a compass needle and a metal wire connected to a battery. Such a wire carries an electric current (see Chapter 12). One of the most startling properties of the interaction was the tendency of the compass to orient itself at right angles to the wire carrying the electric current. Oersted's discovery is the basis of the electromagnet, a magnet consisting of a current-carrying coil of wire that creates a magnetic field. The distributions of iron filings near current-carrying wires are shown in Fig. 3.16.

Electric field. Somewhat less familiar than gravitational or magnetic fields is the electric field, which is the intermediary in the interaction of a hairbrush and the brushed hair. Electric fields also are intermediaries

Figure 3.16 Iron filings near current-carrying wires clearly show the closed loops of the magnetic field lines.



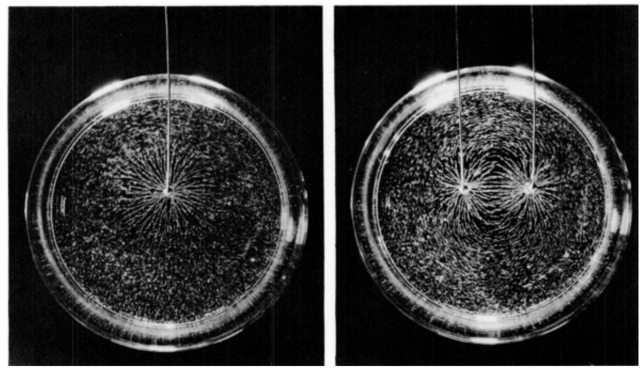


Figure 3.17 Grass seeds suspended in a viscous liquid indicate the direction of the electric field near charged objects.

"... for men still continue in ignorance, and deem that inclination of bodies to amber to be an attraction, and comparable to the magnetic coition ... Nor is this a rare property possessed by one object or two, but evidently belongs to a multitude of objects..."
William Gilbert
De Magnete, 1600

Benjamin Franklin (1706-1790) was born in Boston, Massachusetts. His father was an impoverished candle maker. Ben was apprenticed to an older brother in the printing trade. When his apprenticeship was terminated, he left for Philadelphia where he supported himself as a printer and eventually earned the fortune that freed him for public service. His political career as one of the "founding fathers' of democracy is celebrated, but it is less widely known that he was also one of the foremost scientists of his time. Franklin proposed the one-fluid theory of electricity, and he introduced the terms "positive electricity" and "negative electricity."

in the interaction of thunderclouds that lead to lightning, the interaction of phonograph records with dust, and the interaction of wool skirts with nylon stockings. Objects that are capable of this kind of interaction are called electrically charged.

Electrically charged objects. Objects may be charged electrically by rubbing. Many modern plastic materials, especially vinyl (phonograph records), acetate sheets, and spun plastics (nylon and other man-made fabrics) can be charged very easily. Electric fields originate in electrically charged objects and are intermediaries in their interaction with one another.

Display of the electric field. Individual grass seeds, which are long and slender in shape, like iron filings, orient themselves when they are placed near charged objects (Fig. 3.17). Their ends point toward the charged objects. The patterns formed by the seeds are very similar to the iron filing patterns in a magnetic field. Using the more familiar magnetic field as an analogue model for the electric field, we define the direction of the electric field to be the direction of the grass seeds.

Early experiments with electrically charged objects. Benjamin Franklin and earlier workers conducted many experiments with electrically charged objects. Gilbert had already found that almost all materials would interact with charged objects. One important puzzle was the ability of charged objects to attract some charged objects (light seeds, dry leaves, etc.) but to repel certain others. It was found, for instance, that two rods of ebonite (a form of black hard rubber used for combs and buttons) rubbed with fur repelled one another. The same was true of two glass rods rubbed with silk. But the glass and ebonite rods attracted one another (Fig. 3.18). Since a glass rod interacted differently with a second glass rod from the way it interacted with an ebonite rod, it followed that glass and ebonite must have been charged differently.

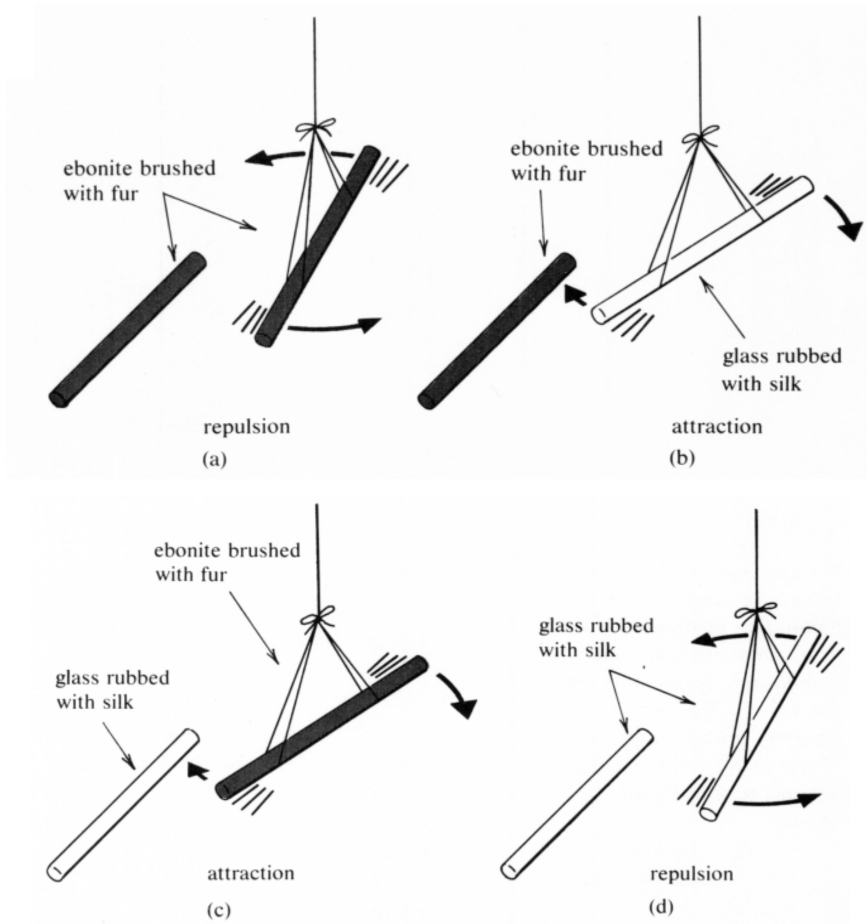


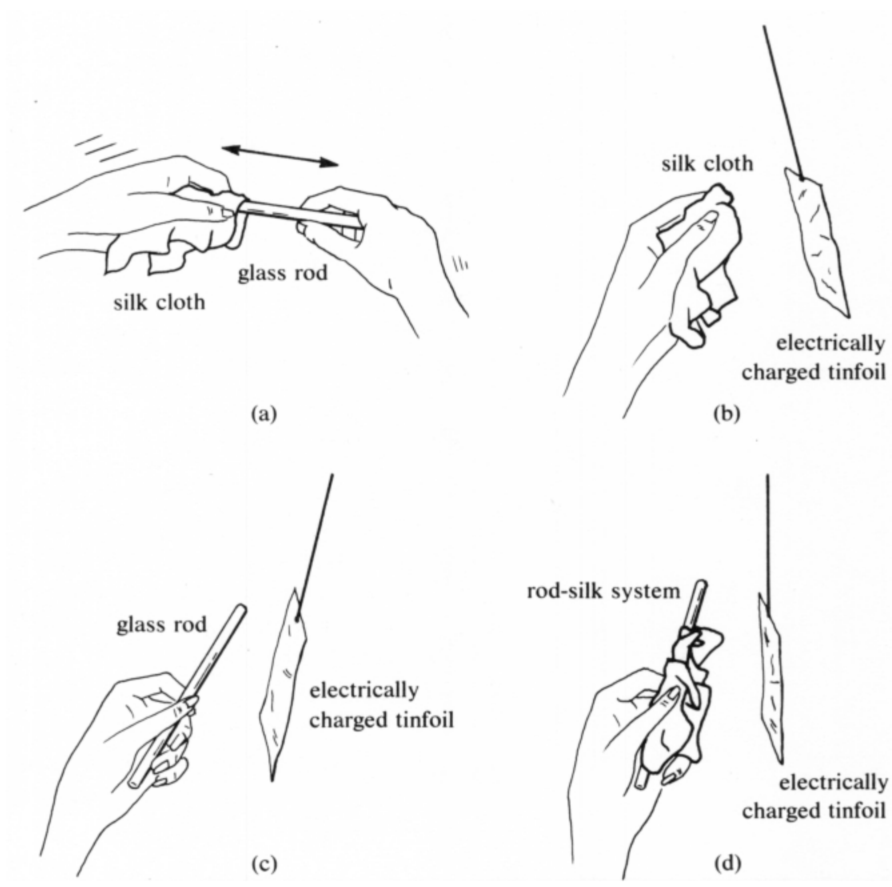
Figure 3.18 Hard rubber rods brushed with fur and glass rods rubbed with silk are permitted to interact. One rod is suspended by a silk thread and is free to rotate. The other rod is brought near until movement gives evidence of interaction.

Two-fluid model for electric charge. Two working models for electrically charged objects were proposed. One model assumed the existence of two different kinds of electric fluids or "charges" (a word for electrical matter) that could be combined with ordinary matter. Charges of one kind repelled charges of the same kind and attracted charges of the other kind.

Franklin's experiment. In a highly original experiment, Franklin found that the two kinds of charges could not be produced separately, but were formed in association with one another. Thus, when an uncharged glass rod and silk cloth are rubbed together, both objects become charged, but with different charges (Fig. 3.19). When the silk is wrapped around the rod, however, the rod-silk system does not have an observable electric field, even though the two objects separately do. Therefore, Franklin concluded that the two kinds of charges were opposites, in that they could neutralize each other. Accordingly, he called them "positive" and "negative," the former on the glass rod, the latter on the silk (and on the ebonite). Combined in equal amounts in one object, positive and negative charges add to zero charge. An object with zero charge is uncharged or electrically neutral.

One-fluid model for electric charge. Franklin's model, to explain this observation, provided for only a single "electric fluid." Uncharged objects have a certain amount of this fluid. Positively charged objects have an excess of the fluid, whereas negatively charged objects have a deficiency. When two uncharged objects are charged by being rubbed together, fluid passes from the one (which becomes negative) to the other (which becomes positive). The fluid is conserved (neither

Figure 3.19 Franklin's experiment.
(a) A glass rod is charged by rubbing a piece of silk.
(b) The silk is tested for electric charge by interaction.
(c) The glass rod is tested for electric charge.
(d) The rod-silk system is tested for electric charge.



created nor destroyed), so the two objects together have just as much fluid as at the beginning of the experiment; hence they form an electrically neutral system. Clearly the isolation of one or two electric fluids would be an exciting success of these models. We will pursue this subject further in Chapters 8 and 12.

Summary

Pieces of matter (objects) that influence or act upon one another are said to interact. The changes that occur in their form, temperature, arrangement, and so on, as a result of the influence or action are evidence of interaction. For the study of interaction, pieces of matter are mentally grouped into systems to help the investigator focus his attention on their identity. As he gathers evidence of interaction, the investigator compares the changes he observes with what would have happened in the absence of interaction. Sometimes he may carry out control experiments to discover this; at other times he may draw on his experience or he may make assumptions.

Pieces of matter that interact without physical contact are interacting-at-a-distance. Radiation and fields have been introduced as working model intermediaries for interaction-at-a-distance. The gravitational field, the magnetic field, and the electric field are the fields important in the macro domain. All three fields have associated with them a direction in space.

List of new terms

interaction	variable factor	field
system	inertia	gravitational field
conservation	inertial mass	magnetic field
state	inertial balance	electric field
evidence of interaction	radiation	magnetic pole
control experiment	interaction-at-a-distance	electric charge

Problems

1. What evidence of interaction might you observe in the following situations? Identify the interacting objects, identify systems (or subsystems) that show evidence of interaction, and describe what would have happened in the absence of interaction.
   (a) A man steps on a banana peel while walking.
   (b) A young man and a young woman pass each other on the sidewalk.
   (c) Two liquids are poured together in a glass.
   (d) A professor lectures to his class.
   (e) A comet passes near the sun.
   (f) Clothes are ironed.

2. Give two examples of interaction where the evidence is very indirect (and possibly unconvincing).
3. Do library research to find what role the interaction concept or other concepts of causation played (a) in Greek philosophy, (b) during the Middle Ages, (c) during the Renaissance, (d) in an Asiatic culture, (e) in a contemporary culture of your choice, and (f) in biblical literature.
4. Interview four or more children (ages 6 to 10 years) to determine their concepts of causation. Raise questions (accompanied by demonstrations, if possible) such as "What makes the piece of wood float?" "What makes the penny sink?" "What makes the clouds move?" "What makes an earthquake?" "What makes rain?" "Can rainfall be brought about or prevented?" (If possible, undertake this project cooperatively with several other students so as to obtain a larger collection of responses.)
5. Give two examples from everyday life of each of the following, and describe why they are appropriate. (Do not repeat the same example for two or more parts of this problem.)
   (a) Systems of interacting objects
   (b) Systems of objects that interact-at-a-distance
   (c) "Control experiments" you have carried out as part of an informal investigation
   (d) Systems that have social inertia
   (e) Systems that have thermal inertia
   (f) Systems that have economic inertia
   (g) Systems that have inertia of motion
6. Analyze three or four common "magic" tricks from the viewpoint of conservation of matter (rabbit in a hat, liquid from an empty glass, etc.).
7. Answer the questions in the margin on p. 59 and explain your answers.
8. Interview four or more children (ages 5 to 8 years) to determine their concept of the conservation of matter. (For suggestions, refer to B. Inhelder and J. Piaget, *The Growth of Logical Thinking from Childhood to Adolescence*, Basic Books, New York, 1958.)
9. Explain how various professions might define systems that are "conserved." Do not use the particular examples in Section 3.3.
10. Compare the use of the word "state" in the phrases "state of a system" and "state of the nation."

11. Propose a systematic series of experiments, other than the one described in Fig. 3.3, to identify the "guilty" circuit breaker.
12. Radiation and fields are introduced as models for intermediaries in interaction-at-a-distance. Describe your present preference for treating interaction-at-a-distance with or without such a model.
13. Four compasses are placed on a piece of cardboard that conceals some small bar magnets. How many magnets are under the cardboard? Locate their poles. Justify your answer. The compass needle directions are shown in Fig. 3.20, below.

bar magnet size: N S

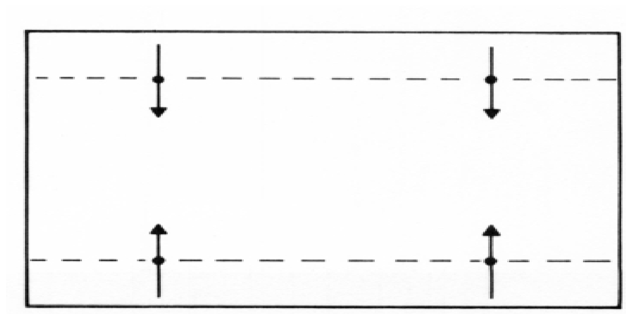


Figure 3.20 Compass needles near concealed magnets (Problem 13).

14. Comment on your present preference for one or the other of the two models described for electrical interaction: the two-fluid model and Franklin's one-fluid model.
15. Describe necessary features of a one- or two-fluid model for magnetized objects. Point out its advantages and disadvantages compared to the pole model.
16. Construct an operational definition for the direction of the electric field.
17. In Section 1.1, "matter" was left as an undefined term. It has been suggested that matter is "anything capable of interaction." Compare this definition with your intuitive concept of matter in the light of Chapter 3, especially Section 3.5. Comment on the logic of this definition, keeping in mind the definition of "interaction." Comment also on the effect of this definition on the conservation of matter principle.
18. Identify one or more explanations or discussions in this chapter that you find inadequate. Describe the general reasons for your dissatisfaction (conclusions contradict your ideas, or steps in the reasoning have been omitted, words or phrases are meaningless, equations are hard to follow, etc.) and pinpoint your criticism as well as you can.

Bibliography

F. Bitter, *Magnets: The Education of a Physicist*, Doubleday (Anchor), Garden City, New York, 1959.

I. B. Cohen, (Ed.), *Benjamin Franklin's Experiments*, Harvard University Press, Cambridge, Massachusetts, 1941.

B. Dibner, *Oersted and the Discovery of Electromagnetism*, Blaisdell, Waltham, Massachusetts, 1963.

W. Gilbert, *De Magnete*, Dover Publications, New York, 1958.

M. B. Hesse, *Forces and Fields*, Littlefield, Totowa, New Jersey, 1961.

E. R. Huggins, *Physics 1*, W. A. Benjamin, New York, 1968. The four basic interactions are discussed in Chapter 11.

E. H. Hutten, *The Ideas of Physics*, Oliver and Boyd, Edinburgh and London, 1967. Chapter 3 is devoted to the field concept.

J. Piaget, *The Child's Conception of Physical Causality*, Littlefield, Totowa, New Jersey, 1960.

J. Piaget, *The Child's Conception of the World*, Littlefield, Totowa, New Jersey, 1960.

J. Piaget and B. Inhelder, *The Growth of Logical Thinking from Childhood to Adolescence*, Basic Books, New York, 1958.

Articles from Scientific American. Some or all of these, plus many others, can be obtained on the Internet at http://www.sciamarchive.org/.

G. Burbridge, "Origin of Cosmic Rays" (August 1966).

A. Cox, G. B. Dalrymple, and R. Doell, "Reversals of the Earth's Magnetic Field" (February 1967).

R. Dulbecco, "The Induction of Cancer by Viruses" (April 1967).

W. P. Lowry, "The Climate of Cities" (August 1967). A discussion of the interactions between human activities and the variables of climate. [Editor's Note: There are many more recent publications on this general topic, with increasing emphasis on "global warming" – the growing evidence that the mean temperature of the atmosphere is increasing and that this is correlated with the use of fossil fuels as well as with an increase in the concentration of carbon dioxide in the atmosphere. For additional information visit the web site of the National Academy of Science (www.nationalacademies.org) and search on "global warming."]

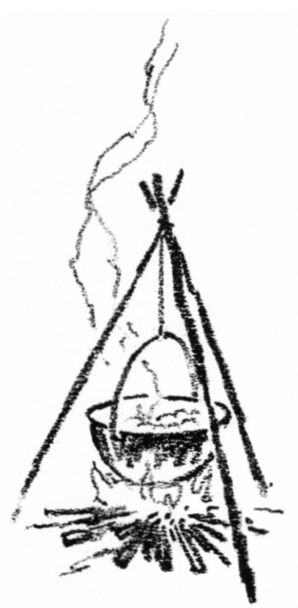
P. J. E. Peebles and D. T. Wilkenson, "The Primeval Fireball" (June 1967).

Matter and energy are of central concern to the physicist. From our ability to make successful theories has come understanding of the ways in which a system may store energy and how energy may be transferred by interaction of objects or systems with one another. From this understanding has come the extensive and effective utilization of energy that is at the base of modern technology and our civilization.

"There is nothing more true in nature than the twin propositions that, 'nothing is produced from nothing' and 'nothing is reduced to nothing' ... the sum total of matter remains unchanged, without increase or diminution."
Francis Bacon
Novum Organum, 1620

Everyone forms qualitative concepts of matter and energy as a result of everyday experience. *Matter* is represented by the solid objects, liquids, and gases in the environment. Matter is tangible; it is capable of interacting with the human sense organs, and various pieces of matter are capable of interacting with one another. Matter appears to be conserved: if an object is once observed in a certain place and later is not there, you are convinced that it has been removed to another location or that it has been made unrecognizable by changes in its appearance. No one believes that it could be annihilated without a trace remaining.

By *energy* is usually meant the inherent power of a material system, such as a person, a flashlight battery, or rocket fuel, to bring about changes in the state of its surroundings or in itself. Some sources of energy are the fuel that is used to heat water, the wound-up spring (or charged battery) that operates a watch, the storage battery in an electric toothbrush, the spinning yo-yo that can climb up its string, the dammed water that drives a hydroelectric plant, and the food that results in the growth of the human body. Your experience with energy is that it appears to be consumed in the operation of the energy sources. Thus, the fuel turns to ashes and becomes useless, the spring unwinds and must be rewound, the battery needs to be recharged, the yo-yo gradually slows down and stops unless the child playing with it pulls properly on the string, and so on. At first glance, therefore, you might conclude that energy, unlike matter, is not conserved.



We will describe two operational definitions of energy in Chapter 9. In the meantime, we will use this term and refer to energy sources, energy receivers, and energy transfer from source to receiver in the expectation that you have an intuitive understanding of these concepts.

4.1 Conservation of energy

Energy transfer. When Sir Edmund Hillary and Tenzing Norgay climbed to the top of Mt. Everest, they expended a great deal of energy. Everyone knows that the two consumed food and breathed air containing oxygen to make this possible. The food, which served them as energy source, came from plants or animals that in turn depended on an energy source (plants or other animals) in a chain of interdependence that is called a food chain. Ultimately, the energy being transferred from organism to organism in the *food chain* can be traced to the sun, which produces energy in the form of light and other radiation.

Energy transfer along the food chain occurs as a result of one

James Prescott Joule (1818-1889) was born near Manchester, the son of a well-to-do brewer whose business he inherited. He devoted himself to science early in life. At 17, he was a pupil of John Dalton, and at 22 he had begun the series of investigations that was to occupy the greater part of his life—the proof that when mechanical energy gives rise to heat, the ratio of energy consumed to heat evolved has a constant and measurable value. Joule's work had supreme significance because it solidly established the principle of the conservation of energy. It was, in Joule's words, "manifestly absurd to suppose that the powers with which God has endowed matter can be destroyed."

organism's eating another, but this is not the only mechanism of energy transfer. You are familiar with other chains of energy transfer. For instance, water escaping from a dam rushes down gigantic pipes to operate turbines, the turbines drive electric generators, and the energy is then distributed by means of transmission lines to factories and residences where some of it may be used to charge the storage battery in an electric toothbrush. The battery finally operates the toothbrush. Depending on the selection of the systems that make up the chain, energy transfer may occur from one system to another (from rushing water to the turbine) or it may occur from one form to another form in the same system (dammed water to rushing water).

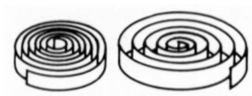
Historical background. During the seventeenth century, many natural philosophers studied rigid-body collisions, such as occur between bowling balls and pins. The bowler transfers energy to the bowling ball, which rolls to the pins and hopefully knocks many of them over, perhaps with so much force that they knock over other pins. The recognition of energy transfer during collisions led Huygens (1629-1695) to a quantitative study from which he concluded that the energy of motion (at that time called *vis viva* or living force, now called *kinetic energy*) was conserved under some conditions. It took many years, however, before scientists realized that the *vis viva* could be transformed into other types of energy and back again with very little loss.

James Watt developed many of the foundations of our modern concept of energy from 1763 on as he invented the steam engine and transformed it from a huge, dangerous monster into the efficient, reliable marvel powering the Industrial Revolution. Basically, the concept of energy provided a way to measure (and helped maximize) the amount of work one could get out of the coal that fueled the early railroad locomotives, textile mills and other factories. Scientists had also simultaneously been discovering the connection between "animal heat" and chemical reactions (metabolism). The third important piece of the puzzle was the recognition that the shaping and drilling of metals by machine tools resulted in a temperature rise.

These three factors inspired the physician Julius Robert Mayer (1814-1878) to speculate on the inter-convertibility of all forms of energy. James Prescott Joule then discovered the quantitative relation between thermal energy and various other forms of energy, establishing a solid experimental foundation for Mayer's theory that is still accepted today.

Since then the *law of conservation of energy* has become generally accepted as one of the most fundamental laws of nature. According to this law, energy may be changed in form but it cannot be created or destroyed. Whenever existing theories have failed to account for all the energy, it has been possible to modify the theory by including new forms of energy that filled the gap. Making these modifications is analogous to your applying the law of conservation of matter to a missing book; you include more locations in your search, and you try to remember who may have borrowed it. You do not believe the book has disappeared from the face of the earth.

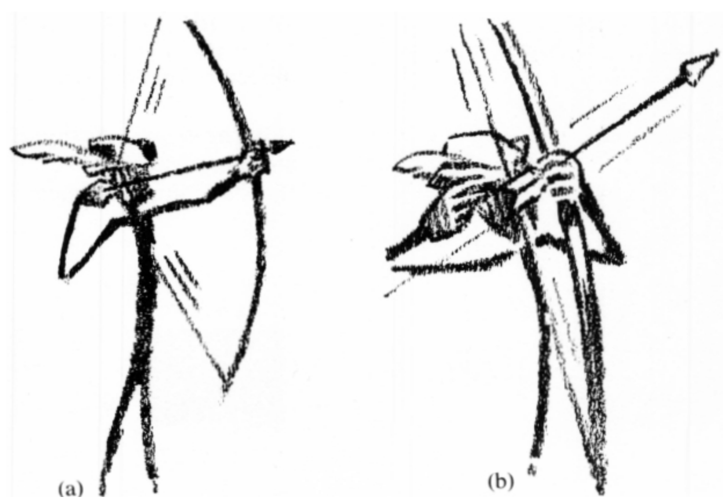
Energy storage. Because energy is conserved, it acquires the same permanence as matter. Just as material objects are kept or stored in certain containers, so we may say that energy is stored in systems (batteries, wound-up springs) that can act as energy sources. During interaction, energy is transferred from the source, which then has less energy, to one or more receivers, which then increase in energy. A system that acted as energy receiver in one process may, during later interaction, act as source and transfer to another system the energy that was temporarily stored in it (Fig. 4.1).



How much energy a system has stored depends on the state of the system. The spring in the state of being tightly coiled has more energy than in the uncoiled state. The storage battery in the charged state has more energy than in the discharged state. We will show in later chapters how the energy stored in a system is related by mathematical models to the variable factors (distance, temperature, speed, and so on) that describe the state of the system.

Energy degradation. These observations suggest that the apparent energy consumption of your everyday experience may actually be only a transfer of energy to a form less easily recognized and less easily transferred further to other systems. During the successive interactions in an energy transfer chain, some energy is transferred to receivers that are difficult to use as energy sources. Examples of these are the warm breath exhaled by Hillary and Tenzing in their climb and the very slightly heated bowling pins that become warm when their motion is stopped by friction with the bowling alley floor or walls. For practical purposes, therefore, their energy is no longer available and appears to have been consumed. In the framework of energy conservation, it is customary to refer to the apparent energy consumption of everyday experience as *energy degradation*.

Figure 4.1 Temporary storage of energy in the bow.
(a) The bow acts as energy receiver.
(b) The bow acts as energy source.



Equation 4.1
energy (initial state) $= E_i$
energy (final state) $= E_f$
energy transfer $= \Delta E$

$$\Delta E = E_f - E_i$$

Identification of systems as energy sources. When you consider the examples mentioned in the introduction to this chapter, you find that we were rather careless in ascribing energy to some of the items mentioned. The food consumed by human beings is not really an energy source capable of sustaining human activity, since it cannot by itself undergo the transformation needed for the release of energy. Instead, the system of food and oxygen is the energy source, whose state can change until the materials have been converted to carbon dioxide and water. This system has more stored energy in its initial state, before digestion and metabolism, than in the final state (Fig. 4.2 and Eq. 4.1).

You see, therefore, that a system must be properly chosen if it is to function as energy source. Just how the system is to be chosen depends on the changes that lead to a decrease or increase of the stored energy. An automobile storage battery functions as the energy source when it is used to operate electrical devices. When the battery is used in a very unusual way—when, for example, it is dropped on walnuts to break them—then the appropriate energy source includes the storage battery, the earth, and their gravitational field. The physicist's definition of systems is geared to ensure conservation of a system (Section 3.3) that functions as energy source or energy receiver. The description of the state of a system includes all the variable factors whose numerical values determine the quantity of energy stored in the system.

4.2 Systems and subsystems

When a system is thought of as energy source or receiver, it becomes worthwhile to choose as simple a system as possible so as to localize the energy. Thus, the spring is only a small part of the watch; the storage battery is only a small part of the electric toothbrush, and so on. This choice of system then will not encompass the entire phenomenon or process being investigated. A system for the entire process will include the energy sources, energy receivers, and other objects participating in the interaction of the source with the receivers. The smaller systems of objects within larger systems are called *subsystems*.

Applications. A simple application of this idea can be made to a slingshot whose rubber band hurls a stone. Both the rubber band and the stone are subsystems of the larger system including slingshot and

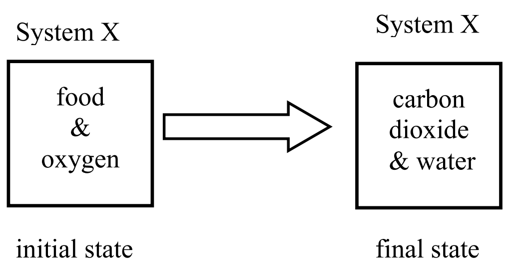
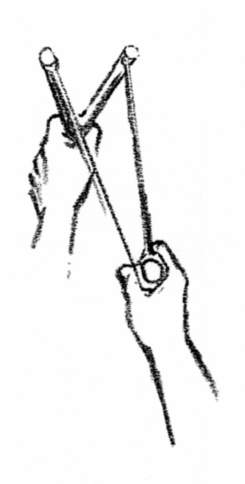


Figure 4.2 Change in the state of System X, which serves as an energy source.



stone. The rubber band acts as energy source and the stone as energy receiver in this instance.

Another application can be made to the South Sea Islander, who starts a fire by twirling a sharp stick against a piece of coco palm bark (Fig. 4.3). A system for the entire process would include the man, the stick, the bow and string, and the piece of bark. The man is the subsystem that acts as energy source. The piece of bark, which gets hot from rubbing, is the subsystem that receives the energy. The stick, bow, and string form another subsystem of interacting objects that facilitate energy transfer.

Selection of subsystems. Before concluding this section, we should add that there are often reasons other than energy transfer for selecting subsystems in a system. For instance, it may be that evidence of interaction is revealed particularly by one subsystem, as when a meat thermometer is used in a large piece of roast beef in the oven. Or, a sample of ocean water may be divided into a water subsystem and a salt subsystem. The subsystems concept is used whenever it is convenient to identify systems that are completely contained within other systems.

4.3 Passive coupling elements

In the slingshot example of the previous section, energy was transferred from the rubber band to the stone by direct interaction. In the making of a fire, however, energy was transferred from one subsystem (man) to the other (bark) through an intermediate subsystem, the bow, string, and stick (Fig. 4.3). This intermediate subsystem never acquires an appreciable amount of energy of its own. A subsystem like the bow, string, and stick, which facilitates energy transfer in a passive way, is called a *passive coupling element*.

Example. There are many situations in which there is a chain of interactions, with energy being transferred from one system to a second,

Figure 4.3 Do the important energy source and receiver in the process interact directly with one another?

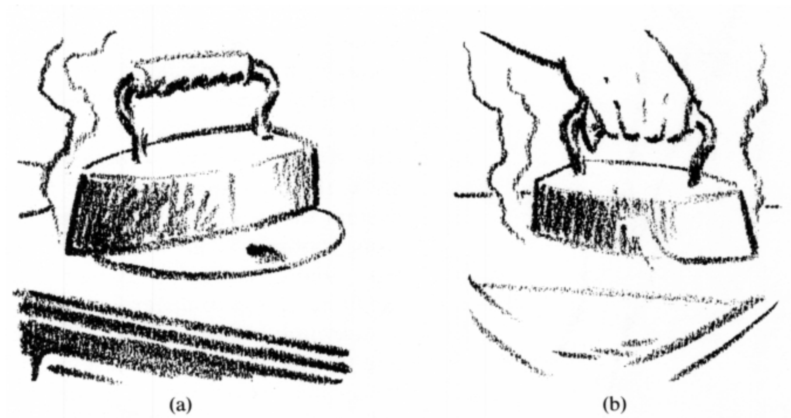


Figure 4.4 Temporary storage of energy in the old-fashioned flatiron. (a) The iron is placed on stove and acts as an energy receiver. (b) The iron acts as an energy source. What is the receiver?

then to a third, and so on. Some of the intermediate systems do not fluctuate in energy but, as it were, pass on the energy they receive as fast as they receive it. A modern electric iron is a good example. Once it has been plugged in for a time and has reached its operating temperature, the energy is supplied from the power line at the same rate as that at which it is passed on to the clothes and the room air surrounding the iron. The energy and temperature of this iron do not fluctuate appreciably. An intermediate system that acts in this way is a passive coupling element. After the iron is unplugged, however, it becomes an energy source and cools off at the same time as it heats the air; then it is no longer only a passive coupling element.

Counter-example. Some of the intermediate systems in an energy transfer chain act as energy receivers and energy sources and are not passive coupling elements; that is, their stored energy fluctuates up and down. An old-fashioned flatiron (which has no electrical heating element) is an example (Fig. 4.4). While it is standing over the fire, it acts as energy receiver and becomes hot. While it is being used to iron clothes, it acts as an energy source and gradually cools off. In other words, its energy content fluctuates along with its temperature.

Passive coupling elements. The concept of a passive coupling element is an idealization that is useful when long chains of energy transferring interactions are to be described and analyzed. The concept can be applied when the energy fluctuation of the intermediate systems are small compared to the energy transmitted. Because the passive coupling elements neither gain nor lose energy, they can be ignored insofar as measurements of energy conservation are concerned. Examples of additional systems that can be treated as passive coupling elements are illustrated in Fig. 4.5.

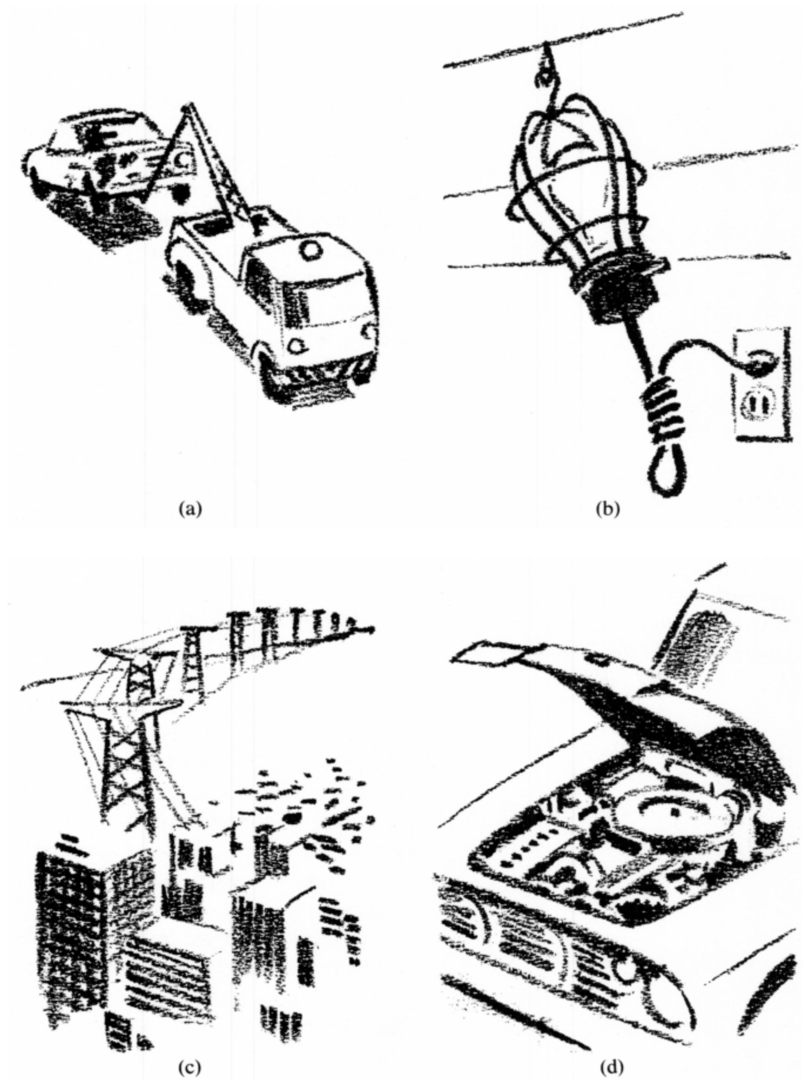


Figure 4.5 Passive coupling elements.
(a) The chain is a passive coupling element between the truck and the smashed car.
(b) The glowing light bulb is a passive coupling element between the power line and the radiated light.
(c) The high-tension line is a passive coupling element between the power station and the city.
(d) An automobile engine is approximately a passive coupling element between the exploding gasoline-oxygen system and the car.

4.4 Forms of energy storage

Consider a rubber band that is stretched and then let fly across a room. This action illustrates an energy transfer, but does not fit into the scheme of subsystems and coupling elements described in the preceding two sections. The energy is always stored in the rubber band, at first because the rubber band is stretched, and after its release because the rubber band is moving relative to the room. To differentiate the different ways in which one system or subsystem may store energy, we introduce the *forms of energy storage*. For the rubber band, we speak of two forms: elastic energy (stretched vs. unstretched rubber band) and kinetic energy (moving vs. stationary rubber band).

The concept of a form of energy storage is useful whether energy is transferred from one form to another within one system (rubber band example) or from one subsystem to another (Section 4.2). The forms of energy stored in a system are associated with the changes that can lead to an increase or decrease of the energy stored in the system. Thus, a pot of water whose temperature may drop has *thermal energy*. The food and oxygen system whose chemical composition may change has *chemical energy*. The steam that may turn to liquid water has *phase energy*. A stretched spring or rubber band that may snap together has *elastic energy*. A moving bullet whose speed can change has *kinetic energy*.

The word "phase" is used in physics with two meanings. The first refers to stages in a repeated motion, as in the "phases of the moon." The second, which we use in this text, refers to "solid," "liquid" and "gaseous phases of matter." That is, phase will refer to the three distinct "states" that matter can take, depending on the temperature and pressure.

Scientists have also discovered a fourth phase, called the "plasma" phase. A plasma occurs at extremely high temperatures, such as at the center of a star, where the particles of a gas break apart into electrically charged particles. Thus a plasma is, essentially, a gas in which each individual particle carries an excess of positive or negative charge. We will not discuss the plasma phase further in this text.

Field energy. In the forms of energy storage just enumerated, the energy was directly associated with one or more concrete objects. This is not the case when you examine systems of objects that interact-at-a-distance. Consider, for instance, the energy that is stored when a stone is raised off the ground. Because of its gravitational interaction with the earth, the stone can fall down and acquire kinetic energy. But from where does the energy come? Neither the stone nor the earth separately is the energy source here; the entire interacting earth-stone system is the energy source.

In Section 3.5 we introduced the concept of the gravitational field as intermediary in the interaction-at-a-distance between the earth and the stone. As the stone is being raised and while it falls, the gravitational field between the two objects changes. Thus, the changes of energy in the earth-stone system are correlated with changes of the gravitational field and not with changes in the stone or the earth separately. We will therefore attribute the energy of the earth-stone system to the gravitational field. In other words, we will take the view that the energy of the raised stone is actually stored in the gravitational field of the earth-stone system. This is an example of a new form of energy storage, *gravitational field energy*.

Electric and magnetic interaction-at-a-distance may be handled in exactly the same way as the gravitational interaction. Thus, a pair of interacting magnets has *magnetic field energy*, and two electrically charged clouds during a thunderstorm have *electric field energy*. The magnetic example is easiest to explore in the macro domain, since you can take two small magnets and manipulate them near each other. Your sensations suggest that the magnets are linked by a spring that pulls them together or pushes them apart (depending on their relative orientation). When you exert yourself, more energy is stored in the magnetic field; when you relax and the magnets spring back, energy is transferred from the magnetic field.

Radiant energy. In Sections 3.4 and 3.5 we described radiation as an intermediary in interaction-at-a-distance. In the example of the candle and the detector, the candle acted as an energy source. When the sun shines on green plants, the sun functions as energy source and the plants as energy receivers. In both cases, radiation carries the energy from the source to the receiver. It is therefore customary to include radiation as a form of energy, called radiant energy.

Radiant energy is similar to field energy in the sense that it is not associated with a material object or system. Nevertheless, it is necessary that you recognize radiant energy if you wish to maintain energy conservation, for there is a time interval after the sun radiates the energy and before the plant receives it. Where is the energy during this time interval? It is not stored in the sun or in the plant; if energy is conserved, it must be stored temporarily as *radiant energy*.

As we pointed out in Section 3.5, the field theory of radiation represents radiation in terms of fields that vary in space and time. In this theory, radiant energy may be classified as field energy. However, since radiant energy manifests itself quite differently from the energy stored in the fields described in the previous subsection, we will refer to radiant energy as a separate form of energy.

Examples. Several examples of how the energy transfer in some common phenomena can be described are illustrated in Fig. 4.6. In these descriptions of energy transfer we have combined the ideas of systems and of forms of energy storage by identifying the forms of energy storage that are important in each system. The idea of a form of energy storage is especially necessary, however, in those examples in which the energy is transferred from one form to another form within the same system or subsystem, as in the rubber band.

4.5 The many-interacting-particles (MIP) model for matter

The atomic model for matter. The question of what happens when matter is subdivided into smaller and smaller pieces has fascinated philosophers for thousands of years. They have also speculated about the existence of a few elementary substances, from which all others were built up. In the eighteenth and nineteenth centuries, the modern science of chemistry was established when the concepts of *element* and *compound* were given operational definitions. According to Lavoisier (1743-1794), a substance was considered to be a chemical element if it could not be decomposed into other substances by any

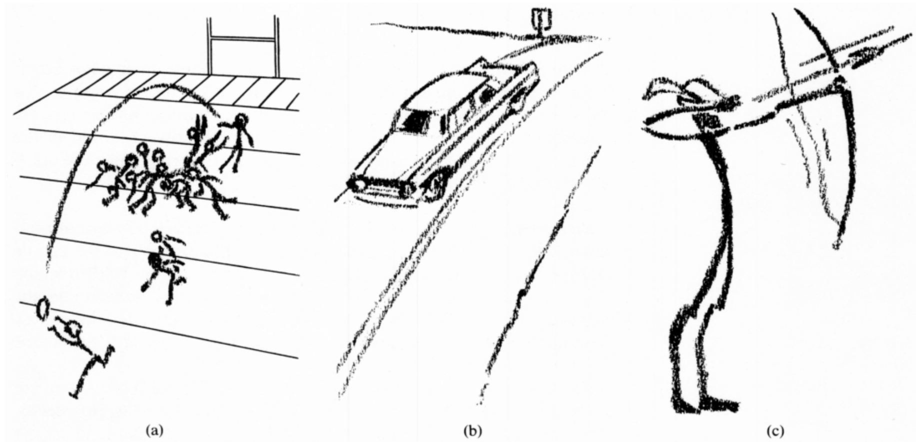


Figure 4.6 Examples of energy storage and transfer (indicates passive coupling elements).*

(a) A long pass in a football game.

System	*Type of energy storage*
passer's arm	*chemical (muscle)*
football (just after throw)	*kinetic*
football & earth system (at top of flight)	*gravitational field (& kinetic)*
football (just before catch)	*kinetic*
receiver's hands & football	*thermal (& kinetic)*

(b) Automobile coasts downhill at a steady speed.

System	*Type of energy storage*
earth & car system	*gravitational field*
*car**	*kinetic*
brake lining	*thermal*

(c) An archer shoots an arrow.

System	*Type of energy storage*
Robin Hood	*chemical (muscle)*
bow	*elastic*
arrow	*kinetic*

means. On the other hand, the substances that could be decomposed were considered to be chemical compounds composed of several elements.

According to this definition, a substance believed to be an element might later be decomposed by some new procedures. It would then be reclassified as a compound. Lavoisier's definition is no longer satisfactory because the development of modern techniques has made it possible to decompose even chemical elements into more primitive components (see Chapter 8). The presently accepted definition of a chemical element is based on properties that are most useful to the chemist.

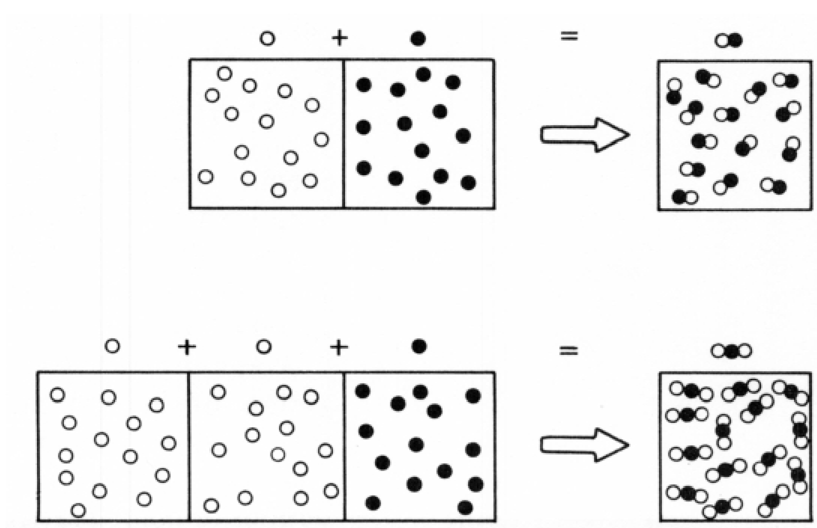
John Dalton (1766-1844) was largely self-taught. He was a retiring person, son of a humble handloom weaver, and from the age of 12 he barely supported himself as a teacher and general tutor in Manchester, England. Dalton possessed strong drive and a rich imagination and was particularly adept at developing mechanical models and forming clear mental images. His astonishing physical intuition permitted him to reach important conclusions despite being only "a coarse experimenter," as his contemporary Humphry Davy called him. Dalton's atomic theory was set forth in A New System of Chemical Philosophy*, published in 1808 and 1810.*

Dmitri Ivanovich Mendeleev (1834-1907) was a professor of chemistry at the University of St. Petersburg in the 1860's when he first noticed that the known elements could by systematized according to chemical properties. His method of classifying and arranging the elements gave us the periodic table, one of chemistry's fundamental conceptual tools. Mendeleev was a compassionate man, deeply involved in the great issues of his time. In 1890, he courageously resigned his chair in protest against the Czarist government's oppression of students and the lack of academic freedom.

Dalton's model. In the nineteenth century it became possible to make a quantitative study of the proportions in which elements combine to form compounds. It was found that elements combine in fixed proportion by weight, and that gaseous elements combine in fixed and small-number ratios by volume. These observations can be explained on the basis of the following working model, based on the ideas of John Dalton and Amadeo Avogadro (1776-1856). Elements are composed of small particles of definite weight called *atoms*; atoms combine in simple numerical ratios to form particles called *molecules*; and every volume of gas under the same conditions of pressure and temperature contains the same number of molecules (Fig. 4.7). This atomic model for matter has been remarkably successful in stimulating chemical research and in accounting, with additional refinement, for the observations made by scientists since Dalton's time.

Existence of atoms and molecules. The atoms and molecules, their interaction, and their motion make up the phenomena in the micro domain described in Section 1.2. Even though atoms and molecules were introduced originally as parts of a working model, they have been so useful in interpreting phenomena in the macro domain that almost everyone now believes that they really exist. Theories based on atoms and molecules have greatly furthered the scientist's understanding of the macro domain. Atoms and molecules are described by many specific and detailed properties, such as mass, size, shape, magnetism, ability to emit light, and so on. Mendeleev found it possible to arrange the elements in a sequence (the Periodic Table) that highlighted similarities in their chemical activity. This sequence was later expanded and slightly revised, and each element was given an atomic number according to its place in the sequence from hydrogen (atomic number 1) to

Figure 4.7 Two chemical reactions according to Dalton and Avogadro's model. Each volume of gas at the same pressure and temperature contains l4 atoms. Atoms combine in simple ratios to form new particles called molecules.



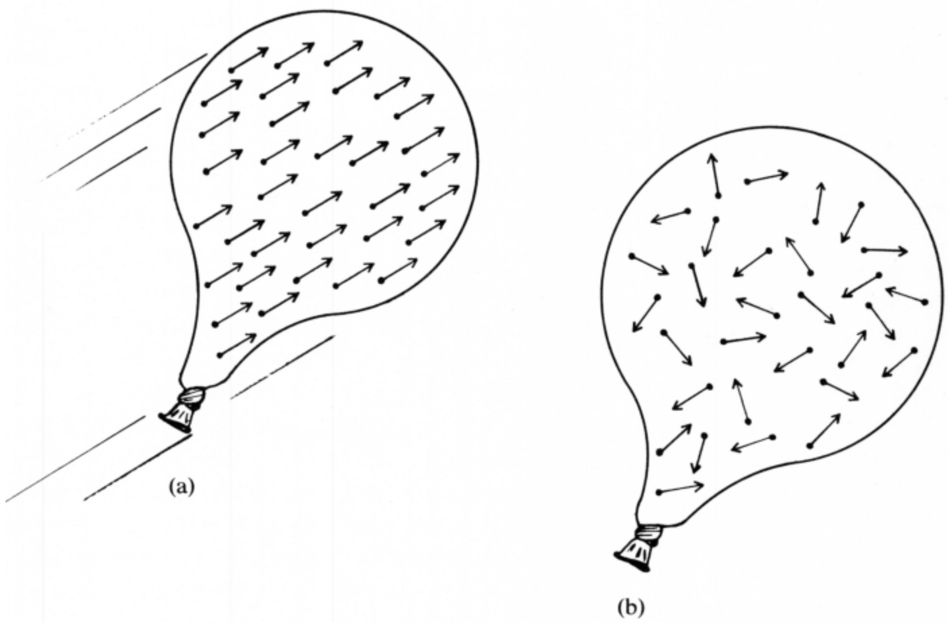


Figure 4.8 Concerted and random action of particles in a balloon.
(a) Macro-domain motion of the balloon is associated with concerted motion of the particles.
(b) Macro-domain failure to move is associated with random motion of the particles.

uranium (atomic number 92). To explain these properties, working models for atoms themselves have been invented; we will describe several of these in Chapter 8.

The MIP model. Many physical phenomena in the macro domain can be explained with the help of a working model in which matter is composed of micro-domain particles in interaction-at-a-distance with one another. The nature of the interactions, the intermediary fields, the sizes of the particles, and other details will be described in Chapter 8 but are not as important as the particles' great number and their ability to interact. We will therefore speak of the many-interacting-particles model (abbreviated MIP model) for matter. We will not identify the particles with atoms or molecules except for associating a different kind of particle with each substance. In the remainder of this section we will describe a few simple examples that illustrate the usefulness of the MIP model. Some of these will be elaborated and made quantitative in later chapters.

Random and concerted action. At the heart of the various applications of the MIP model is the essential idea that the particles are so numerous as to make the action of any single one of little consequence. Instead, only the average action or the most likely combination of actions of many particles is significant in the macro domain. The situation

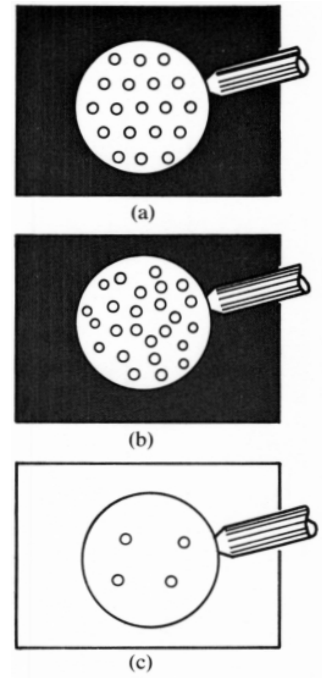


Figure 4.9 MIP model for solids, liquids, and gases.

(a) In a solid material, the particles can only oscillate about their equilibrium positions, which are arranged in a regular pattern.

(b) In a liquid, the particles can move about, but they also interact and are strongly attracted to one another.

(c) In a gas, the particles are so far apart that they rarely interact.

is analogous to that of a human mob, whose members may stampede and act in concert or may be confused and act at cross-purposes. The former is an example of concerted action, the latter of random action.

If, for instance, a blown-up balloon is thrown in a certain direction, then the particles inside the balloon move, on the average, in the same direction (Fig. 4.8a). If, however, the balloon is stationary, the particles move completely at random and their impacts with the rubber skin keep the balloon inflated but do not result in macro-domain motion (Fig. 4.8b).

The MIP model and the phases of matter. The solid, liquid, and gaseous phases of matter in the macro domain can be contrasted particularly easily by use of the MIP model. In a solid material the interacting particles are locked in a rigid pattern, and they can only vibrate a slight distance. In a liquid material the particles are in contact but are not locked into a rigid pattern; in a gas the particles are so widely separated from one another that they hardly interact at all (Fig. 4.9). This explains the rigidity of the solid, the fluidity of the liquid, and the ability of a gas to permeate the entire region of space accessible to it. It also explains the relatively low density of gases compared to solids and liquids (there are large empty spaces between the particles in a gas) and the relatively high compressibility of gases (the particles in a gas can relatively easily be forced into a smaller volume because of the spaces).

The MIP model and mixing. The MIP model also furnishes a ready description of mixing processes where several separate phases in the macro domain combine to form one phase, as when alcohol mixes with water, sugar dissolves in water, or water evaporates. According to the

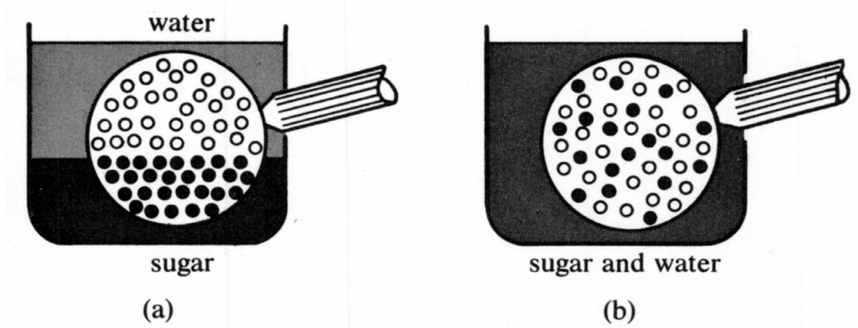


Figure 4.10 Separate sugar and water phases mix to form a single phase.
(a) Separate phases.
(b) Solution phase.

model, the particles of the separate phases mix with one another to produce the uniform solution phase, which includes particles of several kinds. Thus, sugar solution contains sugar and water particles (Fig. 4.10), humid air contains water and "air" particles. (Air is actually a solution containing nitrogen, oxygen, and several other gases, but this fact is unimportant for many physical purposes.) Thus, the interaction of separate phases that leads to solution formation in the macro domain is explained by a mixing of particles in the micro domain.

The MIP model and energy storage. A third use of the MIP model is related to energy storage. In Section 4.4 we identified seven forms of energy storage: kinetic energy, thermal energy, chemical energy, phase energy, elastic energy, field energy (gravitational, electric, magnetic), and radiant energy. These forms of energy storage are appropriate to the macro domain. With the help of the MIP model, thermal energy, chemical energy, phase energy, and elastic energy of systems in the macro domain can be explained as kinetic energy and field energy of the interacting particles (Table 4. l). In the micro domain, therefore, thermal energy, chemical energy, phase energy, and elastic energy do not have separate meanings. The elimination of these four forms of energy storage is an important unifying feature of the MIP model. We will explain how this unification is achieved in later chapters, focusing on the various forms of energy storage.

TABLE 4.1 ENERGY STORAGE ACCORDING TO THE MIP MODEL

Macro domain	*Micro domain*
kinetic energy	*kinetic energy*
thermal energy	*kinetic energy and field energy*
chemical energy	*field energy*
phase energy	*field energy*
elastic energy	*field energy*
field energy	*field energy*
radiant energy	*radiant energy*

The MIP model and interaction. A fourth use of the MIP model is to explain macro-domain interaction through properties of the particles. When two solid objects touch one another (macro domain), then the particles at the surface of one interact with the particles at the surface of the other (micro domain). The interaction that leads to the formation of solutions (macro domain) has already been described as a mixing of the particles (micro domain). When two objects interact-at-a-distance (macro domain), then some or all of the particles in one object interact-at-a-distance with some or all of the particles in the other object (micro domain). A magnet, for instance, is made up of many particles, some of which are magnets; an electrically charged body is made up of many particles, some of which carry an electric charge.

Limitations of the MIP model. Other applications of the MIP model will also be valuable. In all of them, however, it is important for you to remember that the particles of the model are not little metal or plastic balls that roll, bounce, spin, rub, and scrape the way real metal or plastic balls do. You may picture the particles as little balls, but you should be aware that such particles do not have properties beyond the ones assigned to them in the model.

4.6 Equilibrium and steady states

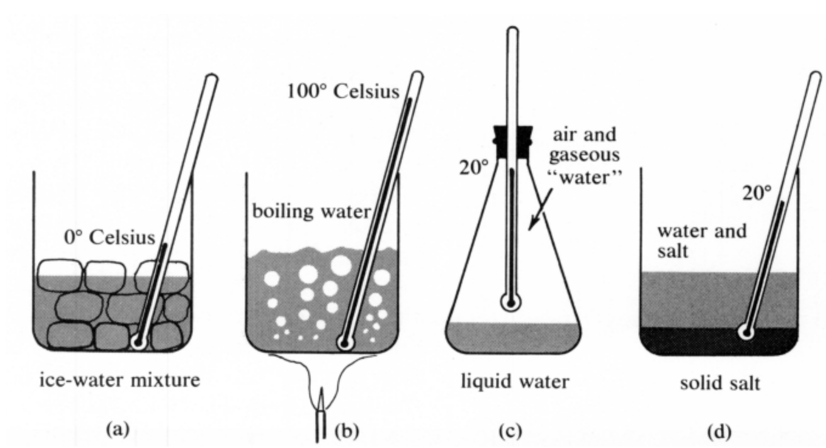
The equilibrium state. It is common experience that systems show changes in their state but that these changes do not continue forever if the system is kept in a uniform environment. A tray of water placed in the freezer, for example, freezes and eventually comes to the temperature of the freezer, but then its state does not change further. A swinging pendulum continues to swing back and forth for some time, but the length of arc of the swing decreases until it comes to rest. A flashlight operates on its battery, but after several hours the light gets dim and finally goes out.

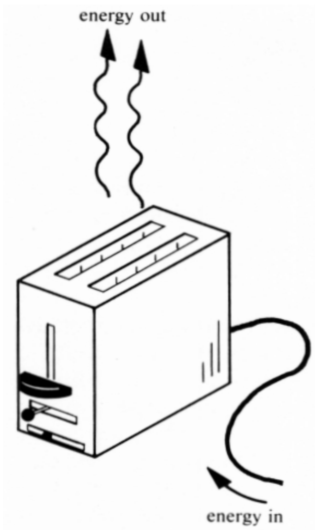
The state of a system that no longer changes in the absence of new environmental interactions is called an *equilibrium state*. A system may come to an equilibrium state in interaction with its environment, such as the water in the freezer, or it may come to equilibrium in isolation from the environment, as did the flashlight. The equilibrium concept is applied to both cases. Since no further change of any kind takes place, no energy transfer occurs either. The equilibrium state of a system, therefore, is a state of maximum energy degradation for that system.

Partial equilibrium. For practical purposes, the equilibrium idea is often applied to changes with respect to only one form of energy storage at a time. In the freezer example, for instance, the equilibrium concept is applied to the temperature of the water, while motion of the water (or ice) is disregarded. In the flashlight example, the charge state of the battery is of interest, and its kinetic or gravitational field energies are not. Therefore, we speak of partial equilibrium, such as mechanical equilibrium (position and motion), thermal equilibrium, chemical equilibrium, and phase equilibrium, whenever equilibrium is reached with respect to the corresponding form of energy storage. Another example is a hot swinging pendulum bob, which may come to mechanical equilibrium (its motion stops) before it comes to thermal equilibrium with the room. Or an ice cube may melt and come to thermal equilibrium with the room air long before its water has evaporated and come to phase equilibrium as a gas mixed with the air. In all these situations, the observer's interest determines which aspect of the system he particularly notes and which details he chooses to overlook.

Phase equilibria. An especially important example of equilibrium states is phase equilibrium, in which two or more phases (solid, liquid, gas) of one substance can coexist indefinitely. At ordinary atmospheric pressure, a pure substance, such as pure water, changes from solid to liquid (melting) or from liquid to solid (freezing) at a certain fixed temperature called the melting temperature. For the substance "water," the melting temperature at atmospheric pressure is 0° Celsius on the internationally accepted scale (formerly called Centigrade) or 32°

Figure 4.11 Phase equilibria.
(a) The equilibrium state of solid and liquid water at atmospheric pressure (reference temperature for Celsius thermometer, 0° Celsius) defines the melting temperature of ice, which is equal to the freezing temperature of water.
(b) The equilibrium state of liquid and gaseous water at atmospheric pressure (reference temperature for Celsius thermometer, 100° Celsius) defines the boiling temperature of water (gaseous water is inside the bubbles).
(c) Equilibrium state of liquid and gaseous water. At room temperature (68°F, 20°C) the gas contains about 2% gaseous water and 98% air.
(d) Equilibrium state of liquid water and salt. At room temperature the liquid contains about 35% salt and 65% water.





How do you tell whether a system is in equilibrium or a steady state?
1) Equilibrium systems do not gain energy from nor lose energy to their environment. Steady state systems steadily gain energy from or lose energy to their environment.
2) An equilibrium system does not tend to change the state of its environment. A steady state system generally does change the state of its environment (possibly gradually).

Fahrenheit. At this temperature, liquid water and solid water (ice) are at equilibrium with one another and can coexist indefinitely. If you attempt to raise the temperature of such a system by supplying heat, the solid water (ice) will all melt before the system begins to get warmer; if you attempt to lower the temperature by removing heat, the liquid water will all solidify before the system gets colder. The melting temperature of pure solid water (ice) has been used to define a reference temperature for thermometer scales, on which it is indicated as 0° Celsius. This and other examples of phase equilibria are illustrated in Fig. 4.11.

The steady state. There are many systems that do not seem to change but that are *not* in equilibrium, for example, the heating coil of an electric toaster, which starts to glow when the toaster is turned on and reaches a steady glow after a few seconds. The state of the coil does not change any more, but the coil is continuously receiving electric energy from the power line and, at the same time, transferring energy at the same rate to the room air. The toaster coil is now acting as a *passive coupling element* between the power line and the room air. Its own state does not fluctuate, and it experiences no net change of its own energy. Such a state, in which a system acts as a passive coupling element transferring energy between other systems, is called a *steady state*. It is to be contrasted with a genuine equilibrium state, in which neither the state of the system nor the state of the system's environment is changing.

An exact steady state is rarely achieved in practice, but it is a useful idealization that is approximated in many practical situations. Even in the toaster example, the room gradually gets warmer and warmer. As a result, the rate of energy transfer from toaster to room is altered, with a consequent gradual change in the "steady" state of the toaster coil. The significance of the steady state is that it represents a balance between energy input and output. In the next section we will describe ways in which a steady state may be maintained, if that is a desirable objective.

Analogue models for equilibrium and steady states. Analogue models for equilibrium and steady states can be created with tanks of water. The water level in a tank represents the state of the system. The approach to equilibrium is modeled by two connected tanks of water (Fig. 4.12), with the connecting valve between them closed. Initially, all the water is stored in the tank on the left. As soon as the valve is opened, water rushes into the second tank and fills the latter up to a level that changes no further. This level represents the equilibrium state.

The steady state is modeled by a different arrangement. Only one tank is supplied with water from a faucet and the water is permitted to escape through a hole near the bottom (Fig. 4.13). Then the water in the tank reaches a level at which the water inflow and outflow are equal. This water level represents the steady state.

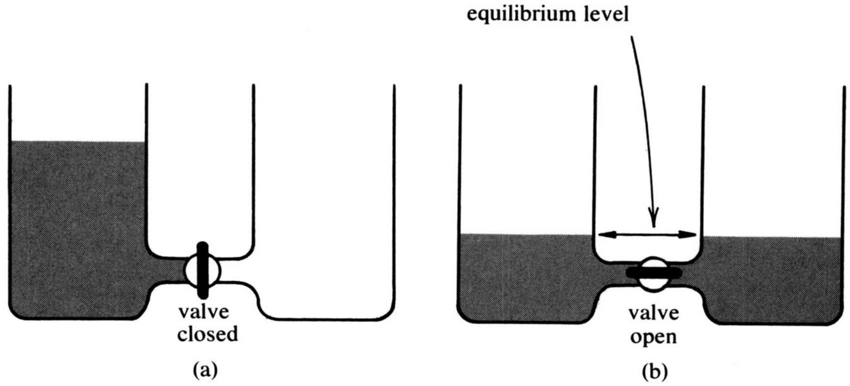


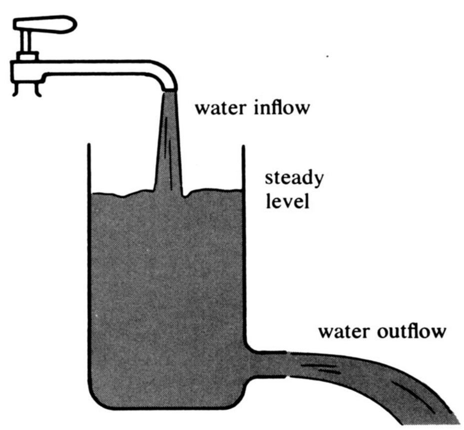
Figure 4.12 Approach to equilibrium is represented by the water level in two connected tanks.
(a) Before interaction.
(b) In equilibrium state with interaction.

Working models for systems in equilibrium and steady states. In studying equilibrium and steady-state conditions, it is frequently useful to construct working models that idealize the actual physical conditions. We took this approach in the example of the toaster, when we assumed that the slowly changing room temperature did not affect the energy output of the toaster wires. Even though real physical systems may never reach equilibrium or steady states, they often come sufficiently close so that a working model, in which certain small interactions or changes in environmental conditions are ignored, does lead to valuable predictions.

The total energy of an isolated system is conserved because the system does not interact and, therefore, does not transfer energy to any other system.

One idealized working model is that of the *isolated system*, which does not interact with its environment at all. An example to which this model may be applied is a system of ice and hot tea interacting with one another in a glass. Treating the ice and tea as an isolated system, which does not interact with the room air, enables you to predict an

Figure 4.13 Steady state is represented by the water level in one tank with inflow and outflow



equilibrium state for the system (the temperature of the iced tea), even though interaction with the glass and the room air means that this state is approached but never reached by the actual iced tea.

A second idealized working model is that of a system interacting with an unchanging environment. Under these conditions the system may come to equilibrium or to a steady state. For example, a real sailboat on a real lake usually encounters rapidly fluctuating wind and wave conditions. Yet a boat designer will first evaluate the boat's performance in mechanical equilibrium—under a steady wind and in the absence of waves—as he determines the size of the sail, length of keel, and so on.

The temperature at any point on the earth shows great day-night and seasonal variations. By the earth's mean temperature we mean the average temperature calculated from all points on the earth's surface at one instant of time.

Applications. An important application of the steady-state concept is to the energy balance of the earth. The earth receives energy from the sun and radiates energy into space. The warmer the earth, the more it radiates. In a working model where the sun is a steady source, the earth's mean temperature will reach a steady value such that the earth's rate of energy loss through radiation is equal to the rate at which it receives energy from the sun.

The steady-state concept has many applications in other sciences as well as in physics. Food supplies remain steady as long as agricultural production is equal to consumption. Excess consumption leads to scarcity and higher prices, excess production to surpluses, lower prices, and loss of farm income. A baby's weight (and yours) increases only as long as the food it eats, water and milk it drinks, and air it inhales exceed losses through elimination, evaporation, and exhaling. When the baby's weight is steady or decreases, there is serious cause for alarm. Deer populations remain steady only as long as the birthrate is equal to the death rate. When deer are protected and their predators are killed, the population is likely to grow until other effects, such as competition for food, result in a new balance. College enrollment increases only as long as more students enter college than are graduating or dropping out. When these two rates come into balance, college enrollments will stabilize.

4.7 The feedback loop model

Stabilization of steady states. In this section we will discuss natural and man-made ways for stabilizing a steady state so that conditions are maintained despite fluctuating external influences. Among the most important natural mechanisms for stabilizing a steady state are the biological systems by means of which warm-blooded animals, including man, regulate their body temperature. A simpler but similar nonbiological example is the regulation of a room's temperature by means of a room thermostat, which controls the operation of a furnace or air conditioner (or both). The thermostat ensures that the room remains at a comfortable temperature level no matter what extremes of temperature may occur outdoors.

Another example of stabilization is furnished by the trained sea lion that balances a ball on its nose by carefully timed movements of its head. Still another illustration of stabilization can be found when you drive a car in heavy traffic and try to maintain a safe distance between your car and the car in front of you in spite of changes in traffic speed. When the space in front of your car widens, your foot goes on the accelerator; when the space narrows, your foot goes on the brake.

These examples illustrate how a condition may be maintained at a steady value even though there are influences that would tend to disturb it. The intermittent operation of the furnace, the head movements of the sea lion, and your braking or acceleration have a stabilizing influence. The *feedback loop model* provides a way to analyze and understand such stabilized situations in which several interactions operate in a loop or circular pattern (Fig. 4.14).

Feedback loop systems. To apply the feedback loop model to a stabilized phenomenon that you observe, you have to take three steps. First, you have to identify the important state or condition that is being stabilized, such as the position of the ball on the sea lion's nose. Second, you have to identify the interaction whereby deviations from the steady state are detected by a system called a *detector* (the sea lion's eyes or the tactile nerves in its nose). Third, you have to identify the interaction that brings about a correction, called *feedback*, to counteract the deviation and restore the desired condition (achieved by the sea lion's head movements). Because the detection and feedback interactions are separate and distinct, the entire regulating process can be diagrammed in loop form (Fig. 4.l5), whence its name is derived.

Negative and positive feedback. The stabilizing influence we described in the specific examples came about because the feedback counteracted the deviation from steady state that triggered the detector. You can conceive of a malfunction of the coupling elements so that the feedback actually makes the deviation worse rather than counteracting it. This situation may lead to catastrophe when, for example, the car in front of you slows down and you step on the gas instead of the brakes.

There are also natural situations where the feedback enhances the deviation rather than counteracting it. When water flows off a mountain, for instance, it erodes the land to form a channel for itself. The channel is the deviation from the previously uniform mountain slope. After further rainfall, more water is gathered in the channel and it flows more rapidly down the steep walls, causing more and more erosion. The process feeds itself until the mountain is deeply eroded. In this example the water combines the roles of detector and "corrector" (by gathering in the previously formed valleys and there concentrating its erosive interaction with the surface soil). Of course, as we said above, the feedback in this example is not corrective but rather increases the deviation.

It is customary to use the term *positive feedback* in situations where the feedback enhances the deviation (erosion example). By contrast, *negative feedback* is used to denote situations where the feedback counteracts the deviation (room thermostat example). Positive feedback, which enhances the deviation, results in catastrophe or brings into play new factors that were omitted from the original feedback loop model.

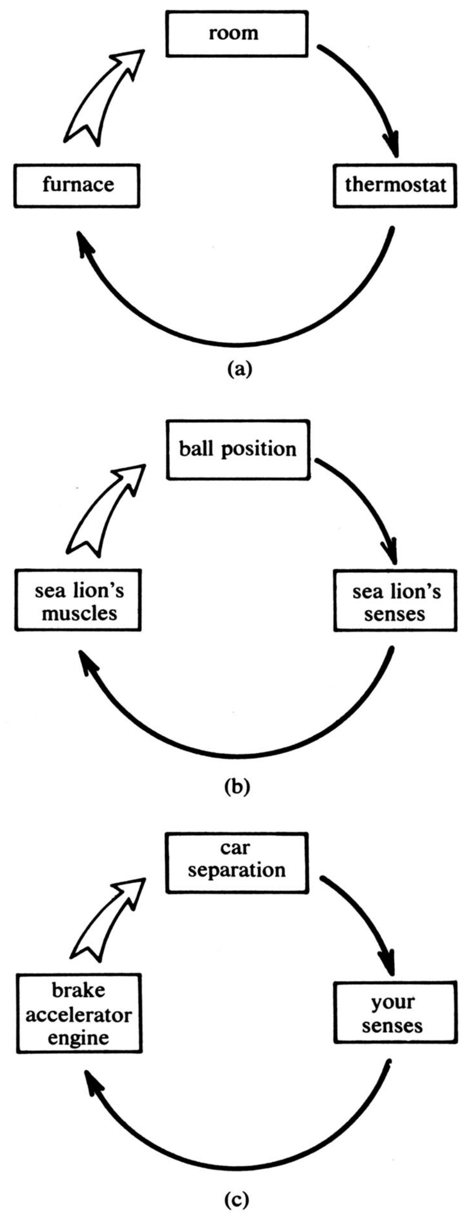


Figure 4.14 Loop or circular pattern of interactions that lead to stabilization of a steady state.
(a) Maintaining room temperature.
(b) Balancing a ball.
(c) Maintaining car separation in traffic.

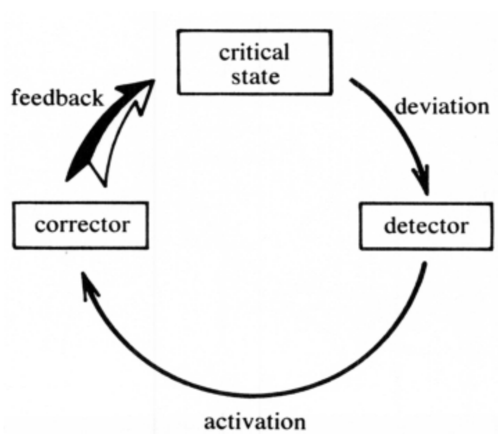


Figure 4.15 The detector identifies a deviation from the critical state. It then activates the corrector, which supplies feedback.

Feedback loop models in the social sciences. Positive and negative feedback are used in the training of animals. Psychologists use the term reinforcement rather than feedback in this context. Nevertheless, the concept is the same. For instance, consider how a pigeon is conditioned to peck at a round shape and not at a square one. Before training, the pigeon does not peck often at any shape. When it deviates from this behavior by pecking at the square shape, a "peck detector" behind the square activates an electric shock, which tends to reestablish the non-pecking behavior (negative feedback). When the pigeon deviates by pecking the round shape, however, it receives food, which tends to enhance this deviation (positive feedback) until the pigeon pecks at the round shape constantly.

Many social phenomena can be analyzed fruitfully with the aid of the feedback loop model. It is important for you to identify the steady condition of the subsystem that is at the focus of your analysis and to describe the deviations that actuate the feedback loop. If the feedback enhances the deviation, it is positive; if the feedback counteracts the deviation, it is negative. Whether positive or negative feedback is more desirable depends on your social values. Negative feedback maintains the status quo, whereas positive feedback leads to evolutionary or even revolutionary change.

Distribution of wealth. In complex phenomena, you must often consider several feedback loops, some with positive and some with negative feedback. Their net effect then depends on the relative effectiveness of the opposing feedback loops. In a social system with a fairly broad distribution of wealth, for example, laissez-faire capitalism seems to provide positive feedback for changes toward a polarized

Figure 4.16 Stabilization of a broad distribution of wealth by a negative and a positive feedback loop.

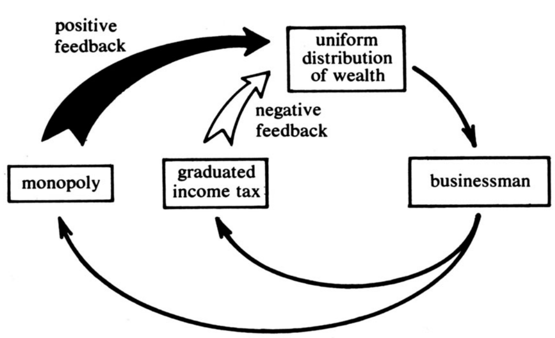
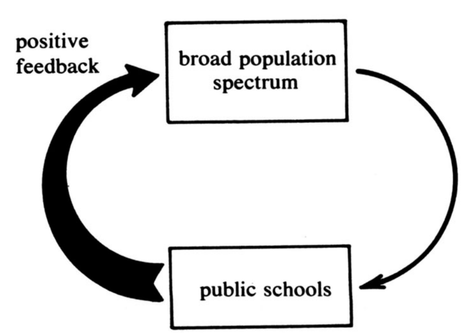


Figure 4.17 Schools can provide positive socioeconomic feedback to the population distribution in a city. When middle-class families leave the city, the schools change (decreased funding, fewer middle-class students) in a way that may repel additional middle-class residents and therefore further reduce the breadth of the population spectrum.



class structure of very rich and very poor individuals. In contrast, a graduated income tax (with higher tax rates for higher incomes) provides negative feedback for such changes and partially stabilizes the original distribution of wealth (Fig. 4.16).

Urban populations. Another currently important social feedback loop leads to the polarization of big city populations into slum dwellers and the very rich. The public schools are one of the important detecting and "correcting" subsystems here (Fig. 4.17). If the middle-class population decreases, the public schools tend to become less attractive to the remaining middle-class residents, who then tend to migrate to the suburbs in still greater numbers. On the other hand, urban schools, if they are funded and organized well enough to offer a superior educational environment, can provide negative feedback and thus contribute to maintaining, or restoring, a broad population spectrum.

4.8 Efficiency of Energy Transfer

When we use energy, we want as much of the energy as possible to go towards achieving our objective. This involves the idea of the *efficiency* of energy transfer. Specifically, e*fficiency* of energy transfer refers to the percentage of the energy from the source that is delivered to the system that we define as the intended receiver, rather than being degraded and/or transferred to other objects. For example, in a toaster, the intended receiver is the bread; the heating of the air is an unintended consequence, and the energy that ends up heating the air can be considered as degraded or "lost." Thus we expect the efficiency of a toaster to be much less than 100%. The toaster itself is a passive coupling element, which simply transfers energy and does not retain any. The total amount of energy is still conserved.

The fact that energy is degraded or considered to be "lost" does *not* mean that the law of conservation of energy is being violated; energy is *always* conserved, and the law of conservation of energy has always turned out to be true. (This is in spite of the stream of people claiming to have invented perpetual motion machines or other ways to generate energy from nothing; in fact, no such claim has ever been substantiated.). In the case of the toaster, a passive coupling element, the total amount of electrical field energy drawn from the source (the electrical utility service) can be measured to be exactly equal to the energy output, that is, the sum of the energy delivered to the toast, air, and other receivers. However, from the point of view of the *user*, the efficiency of a toaster is low because only a small part of the original energy drawn from the power source ended up in the intended receiver (the toast). We will explain efficiency of energy transfer in more detail in Chapter 16.

Summary

The conservation of energy is a powerful physical principle that leads to the concept of energy transfer. Since energy cannot be created or destroyed, all changes in the energy stored by one system must be accompanied by the transfer of energy to or from other systems. The systems that supply the energy are called energy sources, those that receive the energy are called energy receivers, and those that transmit the energy without appreciably increasing or decreasing it are called passive coupling elements.

In addition to being transferred from one system to another, energy can be transformed from one form to another. The major forms of energy storage are kinetic, thermal, chemical, phase, elastic, field, and radiant energy. The many-interacting-particles model for matter helps to explain a large number of properties of matter, including certain forms of energy storage. This model introduces a micro-domain description of physical phenomena to supplement and unify the macro-domain descriptions based on observations made by our sense organs.

The energy stored by a system in equilibrium or steady state does not change. In the former, there is actually no energy transfer between the system and its environment; in the latter there is energy transfer, but no net change because the energy input and output are equal. A system in a steady state, therefore, is a passive coupling element for energy transfer with other systems or the environment. The mechanisms by which a system maintains its steady state are analyzed most effectively by means of the feedback loop model. Negative feedback results in a stable steady state, whereas positive feedback enhances small deviations from a steady state and thereby destroys it. The efficiency of energy transfer is of interest to a user of energy; the efficiency of energy transfer is the percentage of the energy from a source has actually reached its intended receiver or been devoted to a desired end.

List of new terms

matter
energy
energy source
energy receiver
energy transfer
energy conservation
energy storage
energy degradation
subsystem
passive coupling element
kinetic energy
thermal energy
phase energy
chemical energy
elastic energy
field energy
efficiency
radiant energy
atomic model for matter
element
compound
many-interacting-particles model (MIP model)
equilibrium state
steady state
partial equilibrium
phase equilibrium
melting temperature
boiling temperature
isolated system
feedback loop model
positive feedback
negative feedback

Problems

1. Give three examples from everyday life of the apparent degradation of energy.

2. A stage magician performs many amazing tricks. Discuss two or three magic tricks as observations that appear to violate conservation of matter or energy (or both) but really do not.

3. Interview four or more children (ages 8 to 12 years) to explore the meaning they attach to the word "energy" and to the phrase "energy source." Compare their understanding with the modern scientific view.

4. Compare the meanings of the word "law" in the phrases "law of conservation of energy" and "law-abiding citizen."

5. Explain how the law of conservation of energy applies to the examples you gave in response to Problem 1.

6. Give three examples from everyday life of systems that are used to store energy temporarily. Be careful to identify the complete system and compare the state of each system when it stores energy with the state when it has no energy.

7. Consider the following statement about the bow and string in Fig. 4.l: "This intermediate subsystem never acquires an appreciable amount of energy of its own." Explain and criticize this statement. Describe one or two thought experiments that can be used to argue its validity.

8. Discuss (qualitatively) two or more of the examples of coupling elements described in the text with respect to how passive they are (i.e., how much or little of the energy they transfer is stored in them temporarily) and how much the stored energy fluctuates.

9. Give three examples (distinct from those mentioned in the text) from everyday life of systems that usually serve as passive coupling elements. Compare the three examples qualitatively with respect to how passive they are (i.e., how much or little of the energy they transfer is actually stored in them temporarily).

10. Describe how energy is stored and how it is transferred from one form to another in the following examples (see Fig. 4.6). Refer to the forms of energy storage in the macro domain. Begin each description with an energy source.
(a) The ball is kicked off during a football game.
(b) A child hops around on a pogo stick.
(c) A plugged-in electric toaster is used to toast bread.
(d) A pole vaulter vaults over a bar 5 meters high.
(e) An automobile drives up a hill at a steady speed.
(f) A tiger leaps on a mouse.
(g) A photograph is made of a snowy landscape.

11. Interview four or more children (ages 10 to 14 years) to investigate their concept of the structure and phases of matter. Evaluate their responses in relation with the modern scientific view presented in Section 4.5. (Hint: Prepare a few demonstrations where new phases are formed, such as dissolving fruit punch powder and pouring vinegar over baking soda; ask children to explain how the new phases appear from the old ones.)

12. Formulate an acceptable modern definition for "chemical element." You may consult any references you like. Classify the definition as operational or formal and give your reasons.

13. Compare the atomic model for matter with the MIP model described in Section 4.5.

14. Apply the MIP model to illuminate two phenomena from everyday life.

15. Compare the five general applications of the MIP model described in the text and rank them in order from the one that is most meaningful to you to the one that is least meaningful. Explain briefly.

16. Give two examples from everyday life of systems in (approximate) equilibrium with their environment.

17. Give two examples from everyday life of systems in (approximate) steady states.

18. Use the MIP model to construct a micro-domain description of phase equilibrium (e.g., the liquid water, air, and gaseous water in the flask in Fig. 411c).

19. Find an analogue model that clarifies the distinction between equilibrium and steady states. (See the water tank analogue, Figs. 4.12 and 4.13, as an example.)

20. Describe applications of the steady-state concept to two or three phenomena in the social sciences.

21. Describe applications of the equilibrium concept to two or three phenomena in the social sciences.

22. Construct a feedback loop model for:
    (a) the temperature of a refrigerator interior;
    (b) the light intensity reaching the retina of your eyes;
    (c) an example of your choice.
    In each case, identify the system and steady state, describe the feedback mechanism, and explain whether the feedback is positive or negative.

23. Write a critique of one of the feedback loop examples described in the text. Point out the limitations and incorrect conclusions that are implied.

24. The point is made in the text (Fig. 4.17) that public schools could provide positive feedback for changes in the socioeconomic population spectrum of a community.
(a) Apply this model to the conversion of a village into a suburb.
(b) Apply this model to an example that reveals its limitations (i.e. one for which it does not make the correct prediction).

25. Suppose you are an owner of a small store that has a certain annual income from sales as well as expenses such as rent, employees' salaries, purchase of merchandise, and taxes. Whatever is left from the income after expenses is your profit. As the owner, you might think of the percentage of the income that ends up as profit as the efficiency (or profit ratio) of your business. The income is all accounted for; none of the money disappears or is lost, but you consider part of it (the profit) as especially important, and you may want to consider how to increase the efficiency so as to maximize your profit. Compare this situation with the process of energy transfer; explain the similarities with the law of conservation of energy and with the efficiency of energy transfer.

26. Consider the process of energy transfer in an automobile: the original energy is represented by the chemical energy of the fuel; the engine transforms this energy into heat and kinetic energy and thus enables the vehicle to move from place to place. Explain how the law of conservation of energy and the concept of efficiency apply to this process. How would you expect the efficiency to be related to the number of miles per gallon of fuel? How would you expect the efficiency of an SUV and of a racecar to compare with that of a conventional automobile? What might a car designer do to increase the efficiency of a vehicle? How would you expect the efficiency to be related to the expense of running a vehicle? How would efficiency be related to the distance you could travel on a gallon of fuel? How would efficiency be related to the amount of pollution generated by the car? If you wish, find out the efficiency and miles per gallon of various types of vehicles.

25 Identify one or more explanations or discussions in this chapter that you find inadequate. Describe the general reasons for your judgment (conclusions contradict your ideas, steps in the reasoning have been omitted, words or phrases are meaningless, equations are hard to follow, . . .), and make your criticism as specific as you can.

Bibliography

S. W. Angrist and L. G. Hepler, *Order and Chaos, Laws of Energy and Entropy*, Basic Books, New York, 1967.

A. Eddington, *The Nature of the Physical World*, University of Michigan, Ann Arbor, Michigan, 1958.

A. Holden and P. Singer, *Crystals and Crystal Growing*, Doubleday (Anchor), Garden City, New York, 1960.

H. Roscoe, *John Dalton and the Rise of Modern Chemistry*, Macmillan, New York, 1895.

J. Rose, *Automation, Its Anatomy and Physiology*, Oliver and Boyd, Edinburgh and London, 1968. Chapter 2 explores the applications of the feedback concept in automation.

W. Weaver, *Lady Luck: The Theory of Probability*, Doubleday, Garden City, New York, 1963.

N. Wiener, *The Human Use of Human Beings: Cybernetics and Society*,

Doubleday, Garden City, New York, 1954.

Articles from Scientific American. Some or all of these, plus many others, can be obtained on the Internet at http://www.sciamarchive.org/.
S. W. Angrist, "Perpetual Motion Machines" (January 1968).

A. J. Ayer, "Chance" (October 1965).

A. H. Cottrell, "The Nature of Metals" (September 1967).

K. Davis, "Population" (September 1963).

E. S. Deevey, Jr., "The Human Population" (September 1960).

G. Feinberg, "Ordinary Matter" (May 1967).

D. Fender, "Control Mechanisms of the Eye" (July 1964).

M. Gardner, "Can Time Go Backward?" (January 1967).

N. Mott, "The Solid State" (September 1967).

E. J. Opik, "Climate and the Changing Sun" (June 1958).

C. S. Smith, "Materials" (September 1967).

R. Stone, "Mathematics in the Social Sciences" (September 1964).

The nature of light and sound has intrigued mankind over the centuries. Light and sound are connected with sight and hearing and are therefore vital sources of information about our environment as well as essential for survival. The control of sound has led to spoken language and music. The use of light has led to written language and the visual arts. Hearing and seeing, sound and light, enable us to communicate with one another and to derive pleasure from the natural world as well as from music and art.

5.1 Properties of light and sound

Primary sources. We tend to associate light and sound because both are important for sense perception through interaction-at-a-distance. In this way sight and hearing are different from touch, taste, and smell: these three latter senses depend on physical contact of the sense organ and the material (solid, liquid, or gaseous) to be sensed. Both light and sound originate in so-called primary sources: a candle flame, a lightning flash, and the sun emit light; a bowed or plucked violin string, a thunderclap, and a croaking bullfrog produce sound. Both light and sound interact with detector systems: light with the human eye, photographic film, or a video camera; sound with the human ear or a microphone. Both light and sound, therefore, function as intermediaries in interaction-at-a-distance.

Information and energy. Sound and light transmit not only information but also significant amounts of energy. The energy the earth receives from the sun maintains the earth's temperate climate and makes possible the photosynthesis of food material by green plants. The sound blast from dynamite explosions (or Joshua's trumpets at Jericho) breaks windows and may even topple buildings. Even though sound and light always transmit some energy, the amounts involved in seeing and hearing are very small. Therefore it is worthwhile to distinguish between situations in which the transmitted information is more important (such as in sense perception) and those in which the energy transferred is of greater interest (such as in photosynthesis, sunlamps, and dynamite blasts).

Reflection. In describing the similarity of sound and light, we mentioned the existence of primary sources for each. Even though you recognize these primary sources, you are also aware of the reflection of sound and light by all the objects in your environment. As a result of this reflection process, you receive light and sound from all directions and thus appear to be bathed in sound and illumination. Reflecting objects may be called "secondary sources" because they do serve as sources, but their action depends on that of the primary sources.

In other words, the primary sources of light and sound lose energy, but the reflecting objects do not. Rather, the reflecting objects often absorb some of the energy that strikes them and reflect only a part of it. The difference in level of illumination in a room with light-colored walls and in one with dark-colored walls is due to the poor reflection, and greater absorption, of light by the dark walls. A similar effect with respect to sound is achieved by adding rugs and drapes to a bare room. The rugs and drapes are poor sound reflectors (and better absorbers) compared to the bare walls, which reflect most of the sound energy that strikes them.

Sound sources. Sound is generated by vibrating or suddenly moving objects, such as a plucked violin string, a bursting balloon, or a jet engine in operation. Sound travels in air and is blown aside by the wind, but it also travels along a stretched string, through water, and along steel rails. Sound does not propagate in the absence of matter. You can recreate a speaking tube, like the one used by a ship's captain to communicate with the engine room, with the help of an air-filled garden hose. You can also put your ear against a table and listen to the sound produced when you scrape the table surface with your fingernail. Sound is transmitted well by almost all materials, but it is not transmitted well from one material to a very dissimilar material, as from a tuning fork to air.

Can you find evidence that most of the sound is transmitted by the air column inside the hose and not by the walls of the hose?

Loudness and pitch. A sound signal has certain recognizable properties, such as its loudness and, in the case of musical notes, its pitch. The loudness of a sound is related to the energy being transferred from the source to the air and then to the ear. You can reduce the loudness of a sound by weakening the primary source, increasing the distance from the source, or placing obstacles near the source. The pitch of a note is not affected by any of these procedures, but it can be altered by a process called "tuning" of the primary source. A violinist, for instance, tunes his instrument by tightening or loosening the strings. A complex sound, which is a mixture of notes, can be changed by nearby reflecting objects because such objects are usually more efficient reflectors of low notes than of high ones. The sound therefore becomes muffled. You are probably familiar with the effect of a long pipe on the sound of the human voice as well as the fact that the sound of a speaker system depends as much on the room surroundings as on the speakers themselves.

Light sources. Light is usually generated by extremely hot objects, such as a candle flame, an incandescent light bulb filament, or the sun, but it may also come from a fluorescent tube or a firefly. Light travels through air, transparent liquids (like water), and transparent solid materials (like glass), but it also travels through interstellar space, where there is no appreciable amount of material present (vacuum). Most solid objects are not transparent and do not transmit light. They may be white or light-colored and thereby act as efficient reflectors. Photographers have to be especially alert to the color of objects surrounding their photographic subject, because these objects act as "secondary light sources" and influence the exposure conditions. Snow or beach sand may result in overexposure, and dark foliage in underexposure, unless precautions are taken.

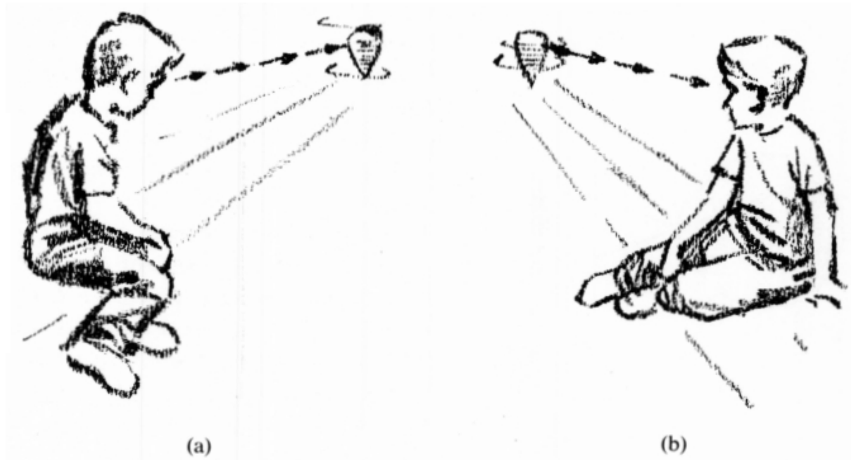


Figure 5.1 Two models for vision.
(a) Child's model that eyes reach out.
(b) Scientific model that light reflected by the object reaches the eye.

Properties of a light signal are intensity and color. As with sound, the intensity is related to the energy being transmitted. You can decrease the perceived intensity of light by weakening the source, going to a greater distance from it, or interposing obstacles. The color depends on the nature of the light of the primary light source, and it can be modified by interposing color filters (green, red, or yellow transparent plastic sheets), or by the presence of strongly colored reflectors whose hue influences the general illumination.

Images. For purposes of sense perception, the most important difference between light and sound is that your eyes give you an incredibly detailed picture (image) of the position, shape, color, and so on, of the objects that reflect light (secondary light sources) in your surroundings, whereas your ears mainly give you information about the primary sound sources. Related to this circumstance is the fact that you possess a primary sound source (your vocal cords) to enable you to communicate with others, but your body is only a reflector and not a source of light. Therefore, your body can be seen only if there is an external source of light and is invisible in the dark. Even when you do control a primary light source, such as automobile headlights, you use it to illuminate and enable you to see *other* objects that reflect light. In other words, you rarely look directly at a primary light source, and you rarely pay attention to secondary sound sources.

Misconceptions about vision. This difference between sound and light, which results in your attending mainly to primary sound sources and secondary light sources, gives rise to a curious misconception among many children. They believe that their eyes (or something from them) "reaches out" to the objects they see, much as their hands reach out to touch objects they wish to feel (Fig. 5.1). In fact, infants must both see and touch objects to develop their visual perception and hand-eye coordination. All of us at one time or other have probably had the feeling that our eyes "reach out." The control we have over selecting the objects we look at and the great detail we can perceive when we look closely, plus the apparent similarity with the way we use our hands, seem to create a strong, but false, impression that the eye is an "active" organ rather than a passive receiving system like the ear. Can you cite any specific evidence that shows that the eye simply receives and senses light, rather than "reaching out"? How might you go about convincing a child who claims his eyes "reach out"?

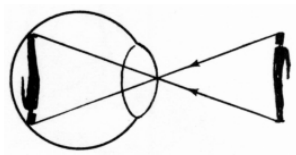


Image on the retina. The sharp visual images are created by the composite effect of many light signals striking the retina of the eye (as shown to the left). Light signals can create an image of a reflecting object because adjacent parts of the object register on adjacent portions of the retina. The retinal image is therefore usually a reliable indicator of the shape of the object. An exception to this rule occurs when the light on its way to your eye passes through a rain-covered automobile windshield or the hot exhaust of a jet engine; then the object appears blurry and you do not register a sharp image.

Ray model for light. Our ability to have visual images, the obstruction of images by objects intervening between object and eye, the silhouette-like shadows cast by objects in bright sunlight, and the projection of film pictures on a screen, all suggest that light usually travels in straight lines. In the ray model, light signals are made up of light rays that travel in straight lines and obey simple geometrical rules as they propagate (move) from source to detector. In the next two sections we will describe details of the ray model and how it may help in the analysis or construction of optical instruments using lenses (eyeglasses, projectors, cameras, microscopes, and telescopes). The ray model, which explains the formation of images and shadows very directly, does not directly explain the existence of a background level of illumination on a cloudy day, when the sun is concealed and there are no sharp shadows. This problem, as well as the problem of the color of light, can nevertheless be solved within the framework of the ray model.

Wave model for sound. It is quite clear that a ray model is not adequate for describing sound because obstacles placed between the source and receiver do not block a sound signal in the same way as they block a light signal. Sound does not appear to travel in straight lines. Instead, it appears to diffuse through space around obstacles much like water waves, which radiate out from a dropped pebble, pass on both sides of an obstacle, and then close again behind it. Sound (and water waves) rarely give rise to a distinct "shadow," that is, a region behind an obstacle from which the effect of the source is completely blocked. This analogy has given rise to the wave model for sound. We will introduce the concept of wave motion in Chapter 6 and apply it to sound in Chapter 7. There you will see how the pitch of a note and the construction of musical instruments are explained by a wave model.

Combination of two sound or light signals. So far we have been describing generally well-known properties of sound and light. We will now describe what happens when light or sound from two sources is combined. The resultant effects of such combinations are surprising

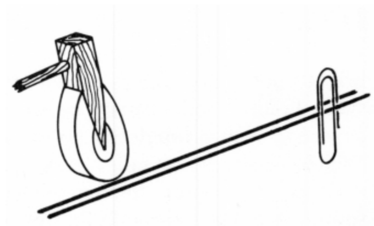


Figure 5.2 Low notes on the piano have two strings. Change the tuning of one string by clipping a paper clip on it. Then hit the key and listen for beats. To compare the sound of two strings with the sound of one, stop the vibration of one string with your finger.

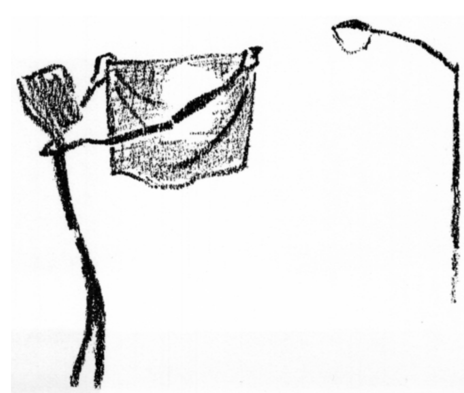
and the implications derived from them, though not widely known, are decisive in formulating working models to explain the phenomena.

Beats. The first phenomenon is the beats that may be heard from a poorly tuned piano. When a single note with two or three strings is struck, there is a rhythmic pulsation of the sound level (so-called beats) even though the tone does not change (Fig. 5.2). To produce beats, it is necessary to have two primary sound sources with pitches that are almost the same. The closer the pitches of the two notes, the slower the beats. The detection of beats is of great importance to the piano tuner, who adjusts the tension of the strings until no beats are heard.

The scientific significance of beats is that there are instants when the combined sound produced by the two sources together (strange as it may seem) is *weaker* than the sound produced by either one alone. This is a very surprising observation, since you would ordinarily expect two similar sources to produce about twice the sound intensity of one alone. You are forced to conclude that the two sources somehow act "in opposition" to one another at the time when the combined sound intensity is very low. To be acceptable, the wave model for sound must therefore make possible this "opposition" as well as the expected reinforcing of two sound sources.

Interference. The second phenomenon involves light. To observe it, look through a handkerchief, a fabric umbrella, or window curtains at a bright, small source of light such as a distant street lamp at night (Fig. 5.3). You will see the light source broken into a regular array of

Figure 5.3 Look through a handkerchief at a bright but distant light. Identify the interference pattern



bright spots. The array of spots will be much larger than the source seen directly, and cannot be explained as arising because the light source is seen through the holes in the fabric. The array of spots may even have touches of color. It is called an *interference pattern*.

An interference pattern is visible whenever one small primary light source is viewed through an obstacle, such as a piece of fabric that has many holes or slits in a regular pattern. Each of the slits functions as a separate "source" and the light passing through each slit "interferes" with the light from all the other slits to produce the observed effect. If you compare the interference patterns caused by coarse and fine fabrics, you will find that the fine fabric creates a larger-scale pattern than the coarse, exactly the opposite from what would happen if you saw the source through the holes in the fabric. With very coarse fabrics (1-millimeter spacing or more) the pattern is too closely spaced to be seen. You can also see rainbow-colored interference effects when you look along a compact disc (CD) or record and let the grooves catch the light, or when you look at thin layers of material in a bright light, such as soap bubbles, mother-of-pearl, or an oil film on water, all of which appear iridescent in sunlight.

5.2 The ray model for light

Many common conditions create the appearance that light rays are visible objects. A powerful searchlight at night is the source of a beam that stabs the sky. A crack in a shutter or curtain admits a shaft of light into a darkened room. Gaps between clouds on a hazy day or trees in mist allow the sun to create streaks of illumination (Fig. 5.4). Actually, light is registered by your eyes only if it strikes the retina. What you see in these examples, therefore, are the massed dust particles or water particles, which reflect the light striking them to your eyes. The form of a beam, shaft, or streak is created by the pattern of the incident light, which is restricted to a slender region of space. The most impressive property of this region of space is that it is straight. Hence the concept of a light ray that travels in straight lines.

Figure 5.4
Streaks of sunlight at dawn.

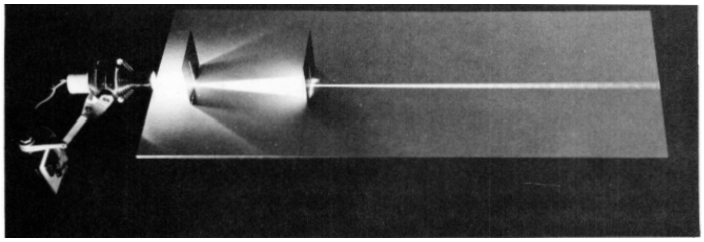


Figure 5.5 A beam and a pencil of light are formed by the small holes in the two screens illuminated from the left.

Figure 5.6 A laser producing a pencil of light.

The phrases "light beam," "pencil of light," and "shaft of light" refer to light traveling in a particular direction, usually directly from a primary source. The distinction among these phrases is a subjective one, with "light beam" referring to the widest lit area and "pencil of light" referring to the narrowest. Their meaning is to be compared with that of "illumination," which refers to light traversing a region of space from all directions, usually coming from reflecting surfaces, so that there are no dark shadows and all objects in the space are visible.

In the ray model, a light beam is represented as a bundle of infinitesimally thin light rays. The width and shape of the bundle determines the area and shape of the beam. The propagation of each ray determines the behavior of the entire beam. In other words, a ray of light is to a beam just as a particle in the MIP model is to a piece of matter (Section 4.5). The function of the ray model is to explain the observations made on light in terms of the assumed properties of light rays. The ray model does not attempt to explain the assumed properties of rays, which are justified (or undermined) by the successes (or failures) of the model.

Properties of light beams. As is true of all models, the observations to be explained by the ray model are built into its assumptions in a simplified and/or generalized form. We therefore begin with the observation of light beams, which the model will have to explain.

Light beams are usually produced by letting light from a powerful source impinge on a screen in which there is a small hole (Fig. 5.5). The light passing through the hole forms the beam. A modern device, the laser (Fig. 5.6), is a primary source that produces a pencil of light.

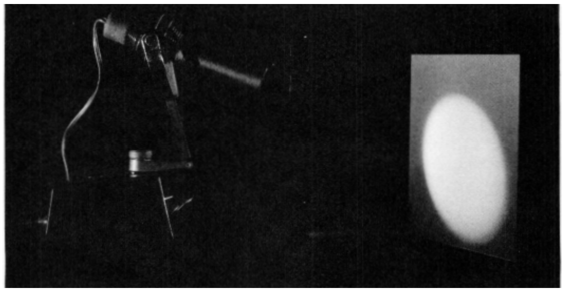


Figure 5.7 The light source and the light spot on the screen can be seen, but not the light beam between them.

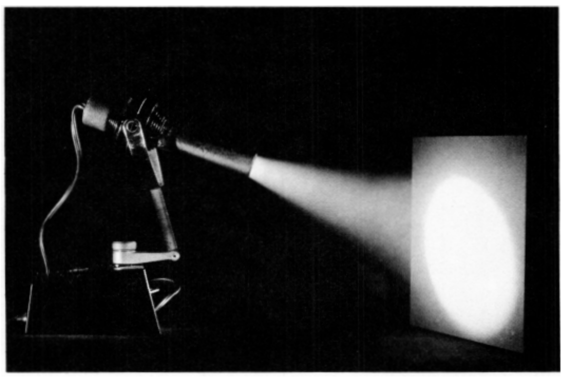


Figure 5.8 Fine smoke particles in the path of the light beam reflect the light and make the beam visible.

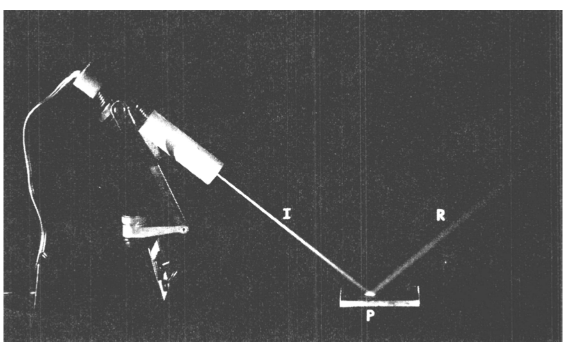


Figure 5.9 A beam of light is reflected from a polished metal mirror (P). Note the incident beam (I) and the reflected beam (R). Why is the mirror mostly dark? Why is the reflected beam dimmer than the incident beam?

If you try this experiment with a flashlight, you wouldn't expect to be able to see the paths of the two light beams as in this photograph. Why not? How might you make the beams visible?

A beam of light cannot be seen from the side, because our eye detects light only when the light strikes our retina (Fig. 5.7). To make the beam "visible," chalk, smoke, or dust particles must be introduced into the region of space (in air or in water) traversed by the beam. The illuminated particles act as secondary light sources and can be seen by reflection, while the un-illuminated particles outside the beam remain dark and unseen (Fig. 5.8). Another technique for tracing a light beam's path is to let it strike a white sheet at a glancing angle (Fig. 5.5).

Three properties of light can be easily observed with beams rendered visible by one of these procedures. 1) A light beam travels in straight lines. 2) A light beam is *reflected* by a polished surface or mirror (Fig. 5.9), and 3) A light beam is partially reflected and partially *refracted*, when it

TABLE 5.1 INDEX OF REFRACTION FOR TRANSPARENT MATERIALS

Material	*Index of refraction*
air	*1.00*
water	*1.33*
glass	*1.5*
diamond	*2.42*
ethyl alcohol	*1.36*

Equation 5.1
(Law of ref<u>le</u>ction)

*angle of ref**le**ction* $= \theta_R$
angle of incidence $= \theta_i$
(Figure 5.11, below, shows θ_R and θ_i .)

$$\theta_R = \theta_i$$

Equation 5.2
(Snell's law of ref<u>ra</u>ction)

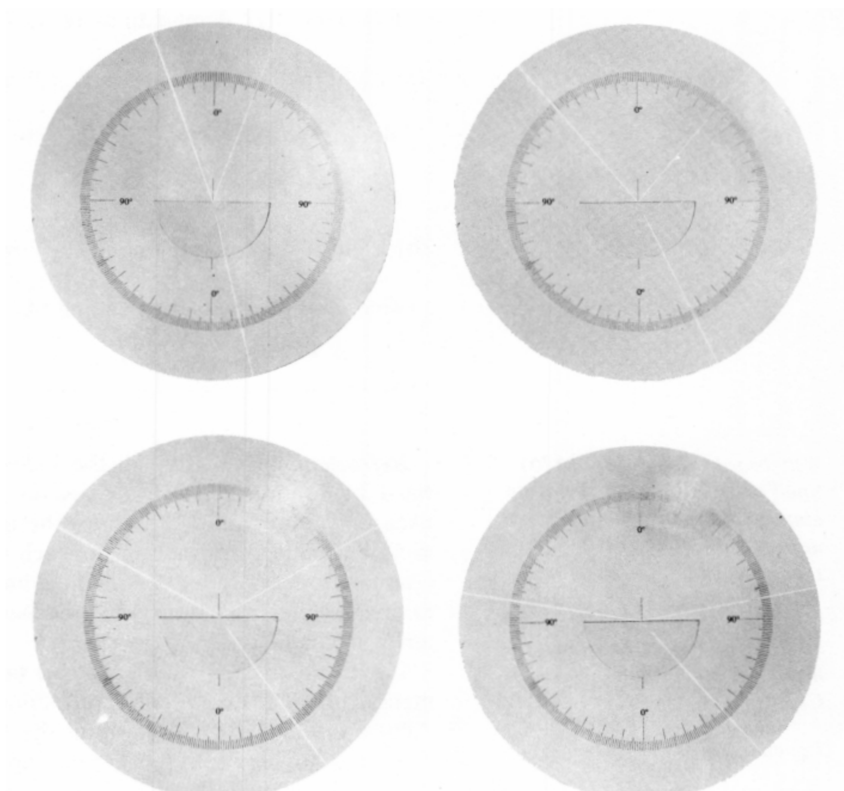
*angle of ref**ra**ction* $= \theta_r$
angle of incidence $= \theta_i$
(Figure 5.11, below, shows θ_r and θ_i .)
*index of ref**ra**ction in material of refraction* $= n_r$
*index of ref**ra**ction in material of incidence* $= n_i$

$$n_r \text{ sine } \theta_r = n_i \text{ sine } \theta_i$$

See Appendix for definition (Eq. A.6) and values (Table A.7) of sine θ.

crosses the boundary between two transparent materials such as air and glass or air and water (Fig. 5.10). Simple mathematical models have been found to fit the observations on the relationships of the angles of reflection, refraction, and incidence (Fig. 5.11, Eqs. 5.1 and 5.2). The symbol "n" in Eq. 5.2 stands for the index of refraction, a number that is determined experimentally for each transparent material (Table 5.1 and Example 5.1).

Figure 5.10 A beam of light enters a semicircular slab of glass at various angles. Note the incident beam, reflected beam, and refracted beam.



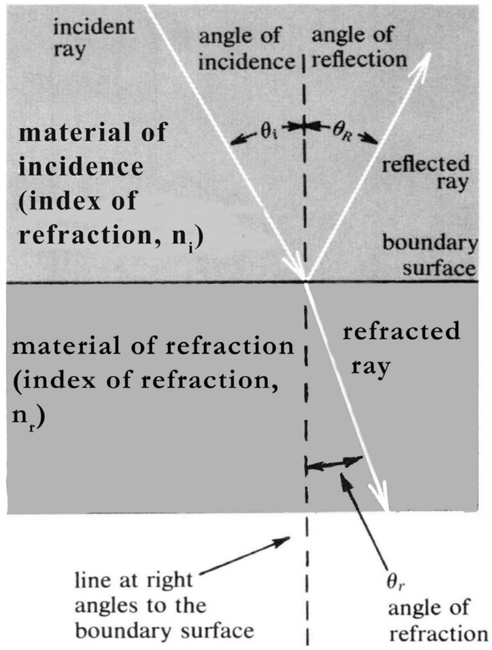


Figure 5.11 The angles of incidence (θ_i), *reflection* (θ_R), *and refraction* (θ_r) *are defined in relation to the line at right angles to the boundary surface between two media.*

Willebrord Snell (1591-1626). Snell's law is an example of an empirical law that summarizes data but does not rest on a theoretical foundation.

Claudius Ptolemy (approx. 140 A. D.) is much better known for his work in astronomy (see Chapter 15).

Equation 5-3 (Ptolemy's Law of Refraction)

$n_r\theta_r = n_i\theta_i$

(less adequate than Snell's Law of Refraction)

EXAMPLE 5-1. Applications of Snell's law.

(a) Light incident on glass: $\theta_i = 40°$, $n_i = 1.00$, $n_r = 1.50$

$$\text{sine } \theta_r = \frac{n_i}{n_r} \text{ sine } \theta_i = \frac{1.00}{1.50} \text{ sine } 40° = 0.67 \times 0.64 = 0.43$$
$$\theta_r = 25°$$

(b) Light incident on diamond: $\theta_i = 60°$, $n_i = 1.00$, $n_r = 2.42$

$$\text{sine } \theta_r = \frac{n_i}{n_r} \text{ sine } \theta_i = \frac{1.00}{2.42} \text{ sine } 60° = 0.41 \times 0.87 = 0.36$$
$$\theta_r = 21°$$

Refraction. Even though the index of refraction of each material must be found, Eq. *5.2* is a useful mathematical model because it uses only one empirical datum (the index of refraction) and yet predicts an angle of refraction for each angle of incidence. This mathematical model is called *Snell's law,* and was formulated by Snell early in the seventeenth century on the basis of experiments conducted with air, water, and glass.

Long before Snell, Ptolemy had tabulated and proposed a mathematical model (Eq. 5.3) for refraction, but the Arabian investigator Alhazen (965-1038) pointed out the inadequacy of Ptolemy's model. The two models are represented graphically in Fig. 5-12, where their similarity for small angles can be recognized. Johannes Kepler (1571-1630), better known for

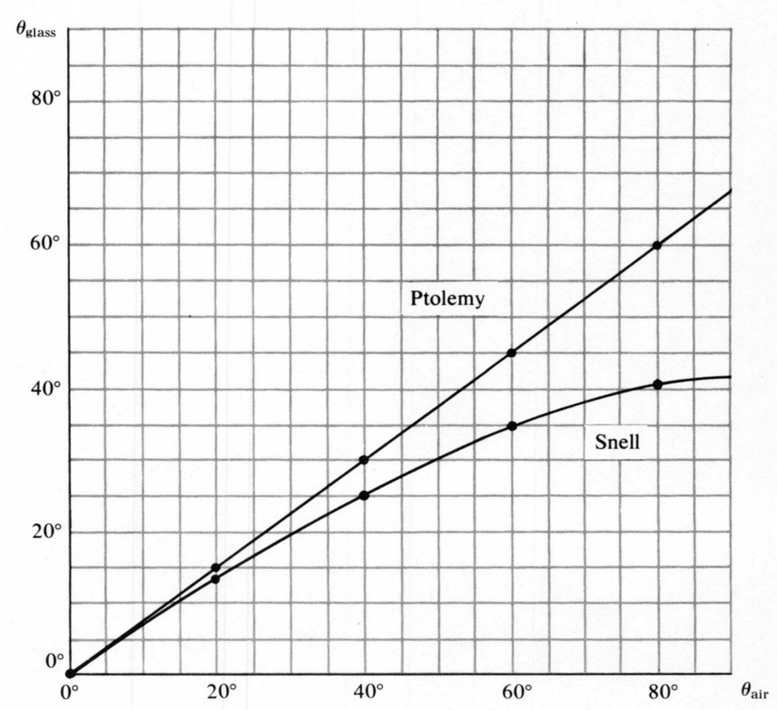


Figure 5.12 Refraction of light passing between air and glass (n_{glass} = 1.5). Ptolemy's model (Eq. 5.3) is compared with Snell's model (Eq. 5.2). For experimental results, see Fig. 5.10. The graph applies for light passing from air into glass or from glass into air. Note that θ_{glass} is always less than θ_{air}.

"I procured me a triangular glass prism, to try therewith the celebrated phenomena of colours... having darkened my chamber, and made a small hole in my window-shuts, to let in a convenient quantity of the sun's light, I placed my prism at its entrance... It was ... a very pleasing divertissement, to view the vivid and intense colours ... I [thereafter] with admiration beheld that all the colours of the prism being made to converge ... reproduced light, entirely and perfectly white, and not at all sensibly differing from the direct light of the sun...."
Isaac Newton
Philosophical Transactions, *1672*

his planetary models for the solar system (see Section 15.2), made measurements of refraction, but was not able to construct a mathematical model better than Ptolemy's (Eq. 5.3). Long after Snell's time, refraction was a key to the acceptance of new models that replaced the ray model for light (Section 7.2).

White and colored light. Isaac Newton achieved a breakthrough in the understanding of light. He found (as others had before) that a glass prism refracted a pencil of light in such a way that a rainbow-colored streak appeared (Fig. 5.13). Newton theorised that the white light was composed of a mixture of various colors. He tested this idea by using a second prism to refract the colored streak back toward its original direction of propagation (Fig. 5.14). The original white light was restored, confirming his theory! When Newton used a screen to isolate only one color in his streak to impinge on the second prism, he found that further refraction of this one color did not alter the color of the light (Fig. 5.15). These findings led Newton to elaborate on the ray model generally accepted in his time to include an explanation of color.

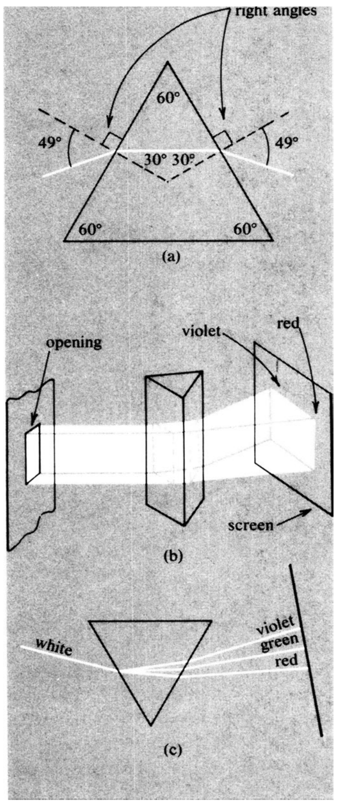


Figure 5.13 (to left) Refraction of light by a glass prism (index of refraction, n = 1.5).
(a) Refraction of a ray of light according to Snell's Law.
(b) and (c) Refraction of a beam of white light. Violet light is bent (or refracted) more than red light.

Figure 5.14 Refraction of a pencil of white light in opposite directions, by two prisms. The colors recombine to make white light. This experiment was first carried out and reported by Isaac Newton.

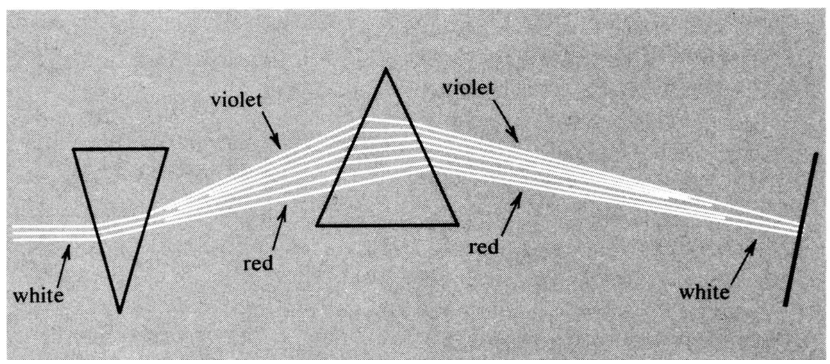
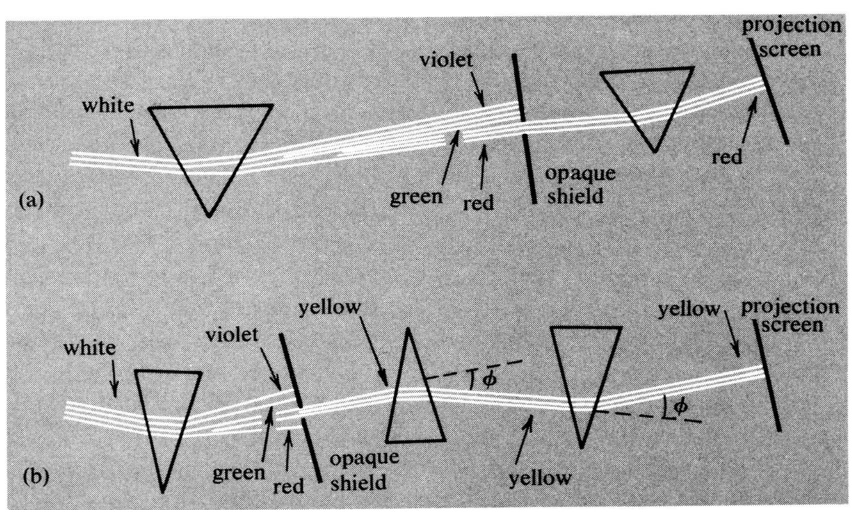


Figure 5.15 Refraction of light of a single color does not change the color. The color is selected from the spectrum by an opening at the appropriate place in the opaque shield.
(a) Red light selected for passage through second prism.
(b) Yellow light selected for passage through two further prisms.



Assumptions of the ray model. Isaac Newton gathered the assumptions of the ray model in his *Opticks, or a Treatise on Reflections, Refractions, Inflections , and Colours of Light*. In his version of the ray model, the elemental light rays were colored (monochromatic = single color), and their color was an intrinsic, unchangeable property. Visual impressions of color were produced by monochromatic rays or by combinations of monochromatic rays, as we will explain below.

The assumptions of the ray model as adapted for this text are summarized in Fig. 5.16 to the right. It is clear that the assumptions were selected so as to be consistent with the observation on light beams described in the previous paragraphs. The task of the model is to explain, in addition, more complicated phenomena, such as the formation of images, the concept of illumination, why not all surfaces act as mirrors, and how colors are produced in mixtures of paint pigments. In all these situations, light is not a single bundle of rays, but a combination of diverging, converging, and crossing rays of various colors and directions of propagation. Assumptions 6 and 7 enable us to apply the model in situations where many rays, following different paths, combine to form an *image* (Section 5.1 and Figure 5.17). Their essence is that only rays that begin by diverging from one point on an object can later converge to form an image of the object point where they started.

Color. The ray model provides three techniques for the production of colored light. One of these is by selective transmission through a color filter. The color filter absorbs rays of some colors and transmits rays of other colors. The colored beam obtained in this way may happen to be monochromatic in Newton's sense, or it may be composed of several monochromatic rays. The second technique is selective reflection by an opaque object. The surface of the object absorbs rays of some color and reflects rays of other colors. The third technique is refraction, through a glass prism or other suitably

Assumption 1. Each point in a primary light source emits rays which diverge in all directions from that point.

Assumption 2. Light rays in a uniform medium travel in straight lines.

Assumption 3. Each point in an object may transmit, absorb, or reflect the rays incident upon it.

Assumption 4. Light rays are reflected, with the angle of incidence equal to the angle of reflection, $\theta_i = \theta_R$.

Assumption 5. Light rays are refracted, with the angles of incidence and refraction of monochromatic rays related by the formula $n_r \sin \theta_r = n_i \sin \theta_i$ *. The index of refraction may differ for different colored rays (dispersion).*

Assumption 6. Whenever rays that diverge from one point in an object meet again at another point on a white screen as the result of reflection and refraction, they make an image of the point in the object.

Assumption 7. An object seen directly, by reflection, or by refraction, appears to be in that place from which the rays diverge as they fall into the observer's eye.

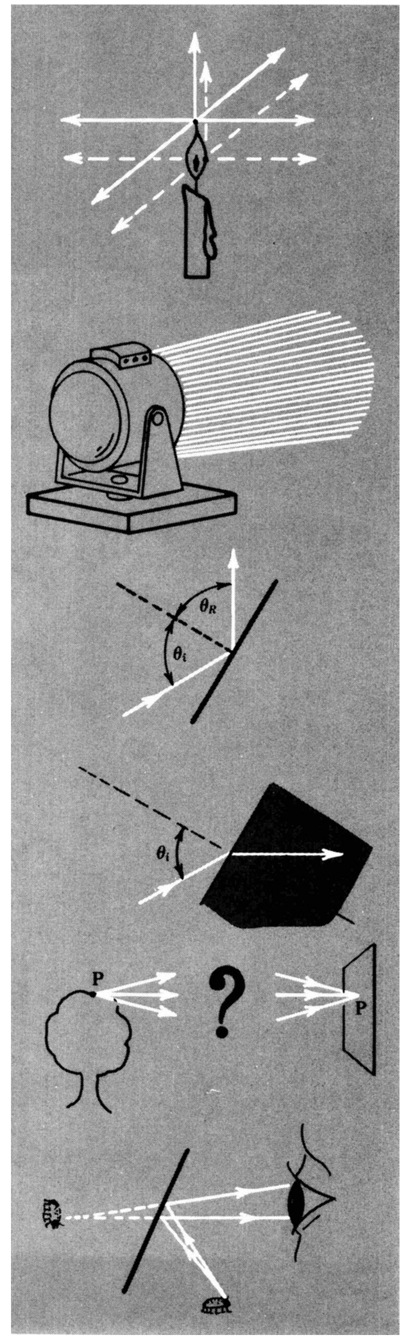


Figure 5.16 Assumptions of the ray model for light.

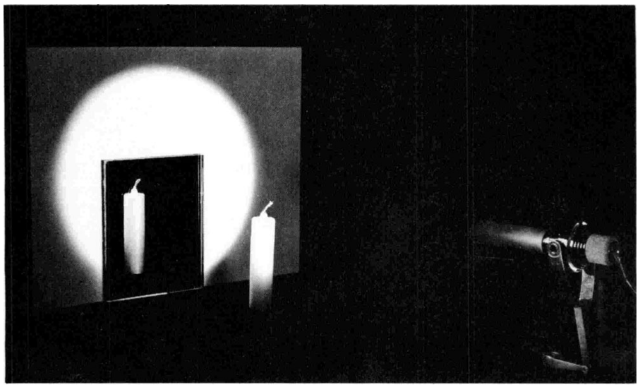
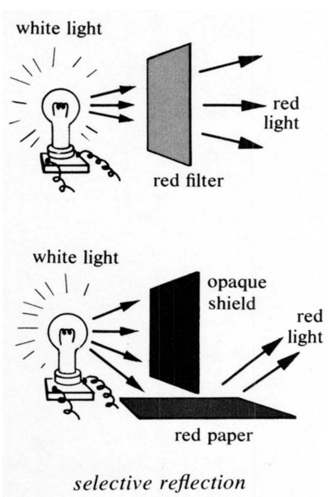


*Figure 5.17 (to left) Specular reflection from a mirror. At first glance this photo may seem unremarkable. However, notice that the mirror is dark and the screen behind it is bright. In addition, both the candle and its image are bright. Finally, notice the light source at the right side: why don't you see it in the mirror? If you can make a ray diagram that shows clearly what is going on here, you understand the ray model very well. (This is **not** in any way a trick photograph; no hidden lenses, mirrors or other apparatus were used.)*

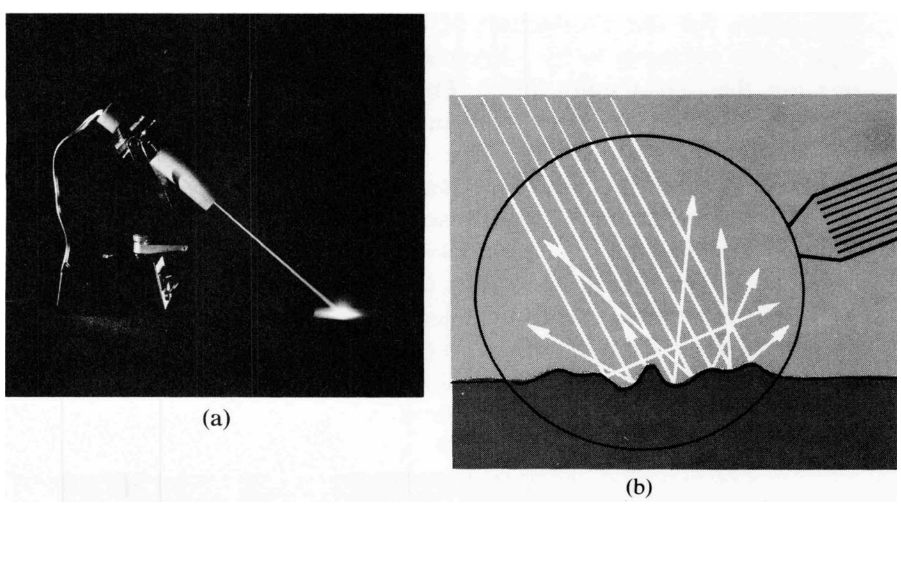


shaped object, such as described earlier (Fig. 5.13). For example, the colors of the rainbow are produced by refraction of light in rainwater droplets. Only refraction is sure to give rise to monochromatic light. Selective transmission and reflection may or may not, depending on the materials in the filter or reflector.

Addition of colors. The ray model can explain the modification of colored light by addition or subtraction of colors. Color addition occurs when several colored beams illuminate the same screen. Where they overlap, the screen acts as a reflecting surface of all the incident colors. This is the process by which Newton obtained white light from the colored streak (Fig. 5.14); after refraction through the second prism, the various colored rays were brought to overlap on the screen.

Subtraction of colors. Color subtraction occurs when filters are inserted into the path of a beam of light, or pigments are mixed in paint.

Figure 5.18 Diffuse reflection.
(a) A pencil of light strikes a dull white surface. Note the diffuse reflected light near surface.
(b) Working model for a dull surface that shows diffuse reflection of light.



The light absorption in the filter removes some colors while others are transmitted. The insertion of successive filters results in the subtraction of more and more colors, until none (that is, no light at all) may be left. Inferences about the absorption by color filters are indirect, because the observer sees only the transmitted light and not what is absorbed (or reflected). To find out what a filter absorbs, you have to illuminate it with monochromatic light and determine whether or not any light at all is transmitted. All the monochromatic light that is not transmitted is absorbed or reflected.

Experiment A. Observations of light absorption by single-color filters

red filter absorbs blue, green, and yellow

yellow filter absorbs blue

green filter absorbs red

blue filter absorbs red and yellow

Experiment B. Observations of light transmission by combinations of single-color filters

red and yellow filters: red light transmitted

red and green filters: no light transmitted

yellow and green filters: yellow and green light transmitted

yellow and blue filters: green light transmitted

On the basis of the ray model and the information in Experiment A above, can you predict the results summarized in Experiment B?

Interference colors. There is a fourth technique that produces colored light: interference. This was mentioned briefly at the end of Section 5.1. It is an observation that cannot be explained by the ray model, since the experiment uses no refraction or selective absorption. Curiously enough, Newton carried out investigations of the color of thin films (air between glass plates, soap bubbles) and tried to adapt the ray model by including multiple reflections and refractions back and forth between the two surfaces of the film. This attempt, however, while suggestive, was not successful and the observations remained unexplained for more than 100 years (Section 7.2).

Models for reflecting surfaces. We have already explained how the color of surfaces arises from selective reflection and absorption of monochromatic light rays. Still to be discussed is the difference between mirrors on the one hand and dull surfaces that reflect light but do not reflect images on the other hand.

Mirror surfaces. A mirror is a surface that maintains the divergence or convergence of rays incident upon it. When you look into a mirror, therefore, you do not see the mirror itself. Rather, you see the light rays reflected from the mirror to your eye, or, more precisely, you see the apparent *sources* of those reflected light rays. (Fig. 5.17). A mirror has a smoothly polished surface to ensure the orderly reflection of all incident rays. This is an example of *specular reflection*. It is well known that fingerprints on a mirror interfere with the specular reflection.

Dull surfaces. A model for reflecting surfaces that are not mirrors must be designed to destroy the divergence or convergence of rays incident upon it. Accordingly, the model surface is highly irregular, with minute irregularities. As light rays strike various points of the surface, they are absorbed or reflected according to Assumption 6. Since the surface is irregular, however, the reflected rays diverge in all directions from a small area on the surface and do not propagate in a direction simply related to the placement of the primary source (Fig. 5.18). This phenomenon is called *diffuse reflection*. According to Assumption 7, an observer will see the surface from which the rays diverge after reflection.

Diffuse reflection, selective reflection and the straight-line propagation of light from primary source to reflecting surface and to the eye account for most of the everyday properties of light and vision. The phenomenon

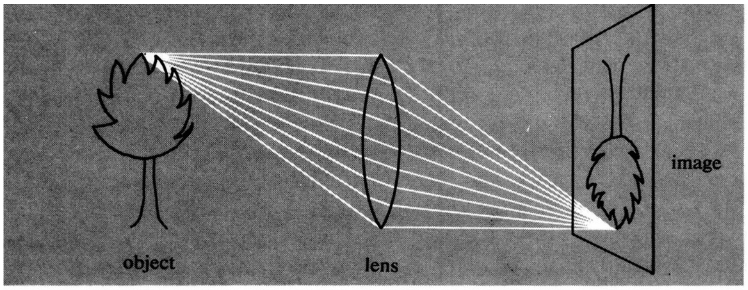


Figure 5.19.
The convex glass lens in air produces an image.

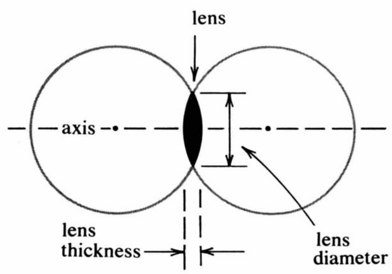


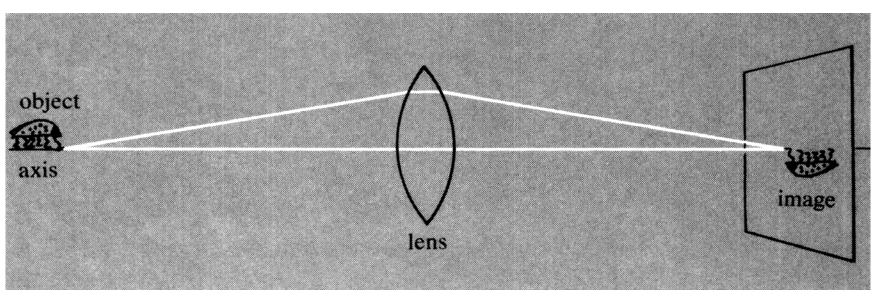
Figure 5.20 A thin spherical lens has the shape of the overlapping region of two slightly interpenetrating spheres.

of refraction enters mainly through the widespread use of lenses in eyeglasses and optical instruments. The next section contains an introduction to the theory of lenses, as derived from the ray model for light.

5.3 Application of the ray model to lenses

Binoculars, telescopes, cameras, and projectors are so common today that it is difficult to believe that none of them existed 400 years ago (and most of them not even 150 years ago). Even eyeglasses were invented only in the fourteenth century, and their use spread very slowly. Until about the time of Galileo the world must have been a blur to

Figure 5.21 Two rays that diverge from an object are refracted by the lens to converge at one point of the image. One ray passes along the axis.



Christian Huygens (1629-1695) was born at The Hague in Holland. His father Constantine, a man of wealth, position, and learning, quickly recognized the boy's unusual capabilities. Christian's father taught him both mathematics and mechanics, and long before his thirtieth birthday, Huygens had published important papers on mathematics, built and improved telescopes, discovered a satellite of Saturn, and invented the pendulum clock. In 1665, King Louis XIV of France invited Huygens to join the brilliant galaxy of intellects that the "Sun King" had clustered about him at Versailles. After 15 years in Paris, Huygens returned to spend his last years in Holland. These last years, however, proved to be as remarkable as his early years. In 1690, Huygens published the Treatise on Light*, his historic statement of the wave theory of light.*

Waves on a water surface are such a familiar and expected occurrence that a completely still, glassy pool excites surprise and admiration (Fig. 6.1). You can also observe waves on flags being blown by a strong wind. In this chapter you will be concerned with how waves propagate, what properties are used to describe them, and how waves combine with one another when several pass through the same point in space at the same time. In the wave theory, which was formulated by Christian Huygens during the seventeenth century, the space and time distribution of waves is derived from two assumptions, the superposition principle and Huygens' Principle. The wave theory is very "economical" in the sense that far-reaching consequences follow from only these two assumptions.

Waves are important in physics because they have been used in the construction of very successful working models for radiation of all kinds. You can easily imagine that dropping a pebble into a pond and watching the ripples spread out to the bank suggests interaction-at-a-distance between the pebble and the bank. The waves are the intermediary in this interaction, just as radiation was the intermediary in some of the experiments described in Sections 3.4 and 3.5. In Chapter 7, we will describe wave models for sound and light and how these models can explain the phenomena surveyed in Chapter 5. The success of these models confirms Huygens' insight into the value of wave theory. However, Huygen's contributions and wave theory were not fully appreciated and exploited until the nineteenth century.

Waves were originally introduced as oscillatory disturbances of a material (called the *medium*) from its equilibrium state. Water waves and waves on a stretched string, the end of which is moved rapidly up and down, are examples of such disturbances. The waves are emitted by a source (the pebble thrown into the pond), they propagate through the medium, and they are absorbed by a receiver (the bank). Even though waves are visualized as disturbances in a medium, their use in certain theories nowadays has done away with the material medium. The waves in these applications are fluctuations of electric, magnetic, or

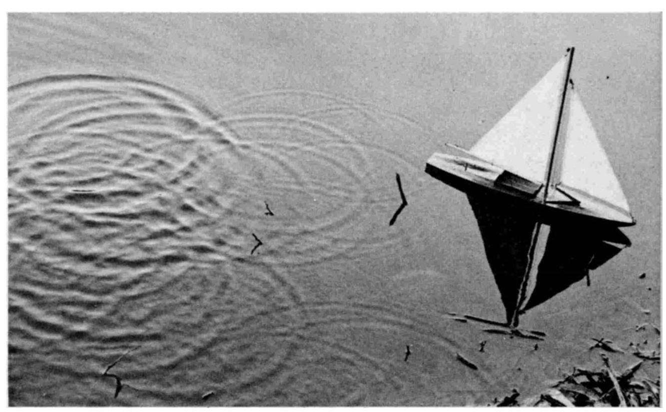


Figure 6.1 The reflected image gives information about the smoothness of the water surface. Why are the reflections of the sails dark and not white?

gravitational fields, rather than oscillations of a medium. The use of such waves to represent radiation has unified the radiation model and the field model for interaction-at-a-distance (Section 3.5). Our discussion here, however, will be of waves in a medium and not of waves in a field.

6.1 The description of wave trains and pulses

Oscillator model. We will analyze the motion of the medium through which a wave travels by making a working model in which the medium is composed of many interacting systems in a row. Each system is capable of moving back and forth like an oscillator, such as the inertial balance shown below and described in Section 3.4. You may think of the oscillators in a solid material as being the particles in an MIP model for the material.

Amplitude and frequency. Each oscillator making up the medium has an equilibrium position, which it occupies in the absence of a wave. When an oscillator is set into motion, it swings back and forth about the equilibrium position. The motion is described by an *amplitude* and a *frequency* (Fig. 6.2). The amplitude is the maximum distance of the oscillator from its equilibrium position. The frequency is the number of complete oscillations carried out by the oscillator in 1 second.

Interaction among oscillators. When waves propagate through the medium, oscillators are displaced from the equilibrium positions and are set in motion. The wave propagates because the oscillators interact with one another, so that the displacement of one influences the motion of the neighboring ones, and so on. Each oscillator moves with a frequency and an amplitude. It is therefore customary in this model to identify the frequency and amplitude of the oscillators with the frequency and amplitude of the wave. In addition, as you will see, there are properties of the wave that are not possessed by a single oscillator but that are associated with the whole pattern of displacements of the oscillators.

Conditions for wave motion. The oscillator model described above has two general properties that enable waves to propagate. One is that the individual oscillator systems interact with one another, so that a displacement of one influences the motion of its neighbors. The second is that each individual oscillator has inertia. That is, once it has been set in motion it continues to move until interaction with a neighbor slows it down and reverses its motion. These two conditions, interaction and inertia, are necessary for wave motion.

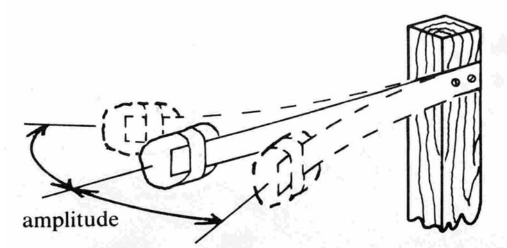


Figure 6·2 An oscillator in motion.

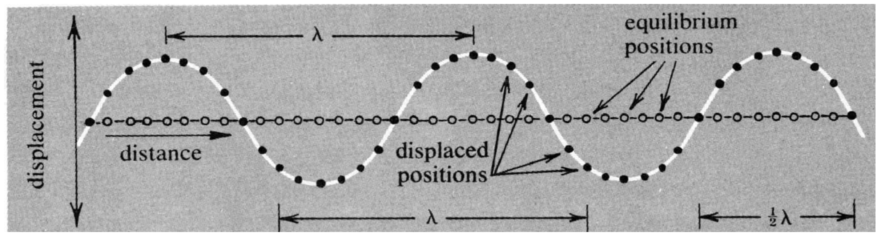


Figure 6.3 Row of oscillators in a medium, showing equilibrium positions and displaced positions in a wave. The wavelength is the distance after which the wave pattern repeats itself.

Equation 6.1

wavelength (meters) = λ

wave number (per meter) $= k$

$$\lambda \times k = 1,\ k = \frac{1}{\lambda},\ \lambda = \frac{1}{k}$$

EXAMPLES

$\lambda = 0.25$ m

$$k = \frac{1}{\lambda} = \frac{1}{0.25\,m} = 4/m$$

This is 4 wavelengths/m.

$\lambda = 5.0$ m

$$k = \frac{1}{\lambda} = \frac{1}{5.0\,m} = 0.2/m$$

This is 0.2 wavelengths/m.

$\lambda = 0.0001$ m $= 10^{-4}$ m

$$k = \frac{1}{\lambda} = \frac{1}{10^{-4}\,m} = 10^4/m$$

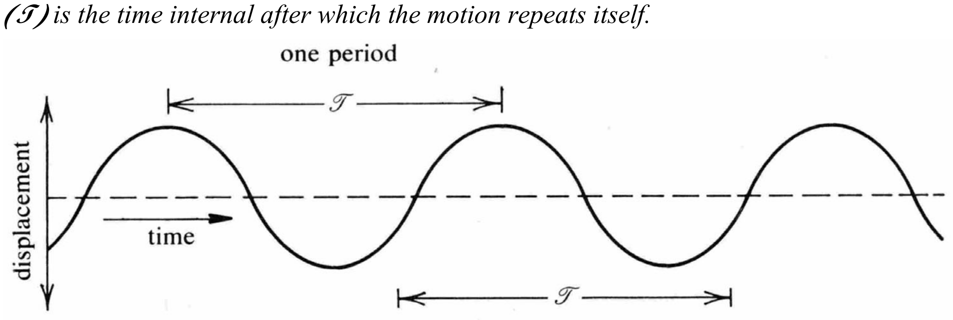
This is 10^4 or $10{,}000$ wavelengths/m.

Wave trains. Look more closely now at the pattern of the oscillators in the medium shown in Fig. 6.3. As a wave travels through the medium, the various oscillators have different displacements at any one instant of time. The wave is represented graphically by drawing a curved line through the displaced positions of all the oscillators (shown above in Fig. 6.3). This curved line, of course, changes as time goes on because the oscillators move. Note, however, that the individual oscillators in the model move only up and down.

Wavelength and wave number. You can see from Fig. 6.3 that the wave repeats itself in the medium. This pattern of oscillators is called a wave train, because it consists of a long train of waves in succession. A complete repetition of the pattern occupies a certain distance, after which the pattern repeats. This distance is called the *wavelength*; it is measured in units of length and is denoted by the Greek letter lambda, **λ**. Sometimes it is more convenient to refer to the number of waves in one unit of length; this quantity is called the *wave number* and it is denoted by the letter **k**. Wavelength and wave number are reciprocals of one another (Eq. 6.1).

Period and frequency. We have just described the appearance of the medium at a particular instant of time. What happens to one oscillator as time passes? It moves back and forth through the equilibrium position as described by a graph of displacement vs. time (Fig. 6.4) that is very similar to Fig. 6.3. The motion is repeated; each complete cycle requires a time interval called the *period* of the motion, denoted by a script "tee," $\mathcal{T}$. The number of repetitions per second is the

Figure 6.4 Graph of the motion (displacement) of one oscillator over time. The period ($\mathcal{T}$) is the time internal after which the motion repeats itself.



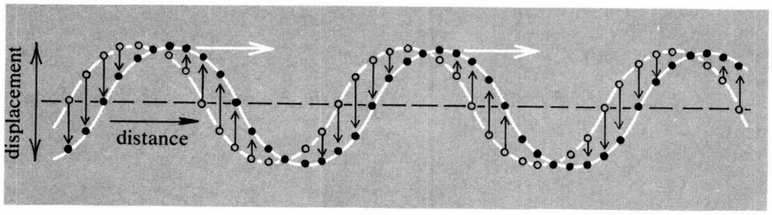
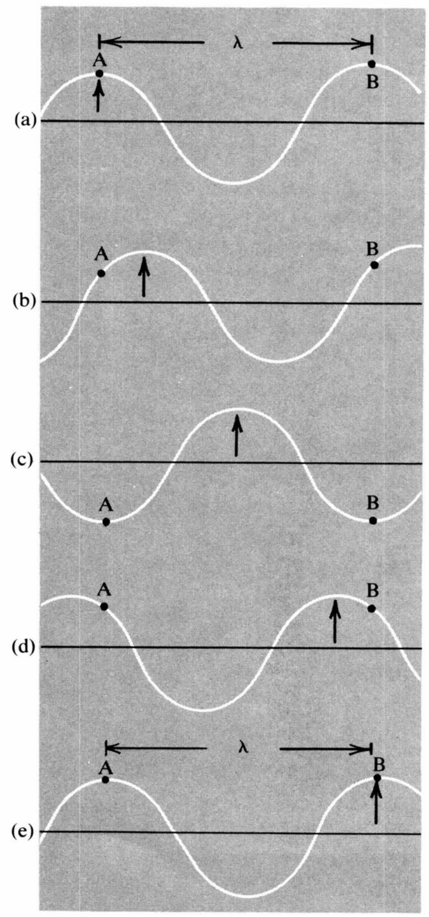


Figure 6·5 The wave moves to the right as the oscillators move up and down. The black circles and black dots represent the displacements of the oscillators at two different times.

Figure 6·6 In one period, oscillators A and B carry out a full cycle of motion from crest to trough and to crest again. The crest initially at A moves to B in this time interval.



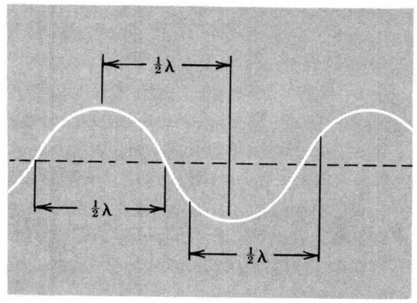
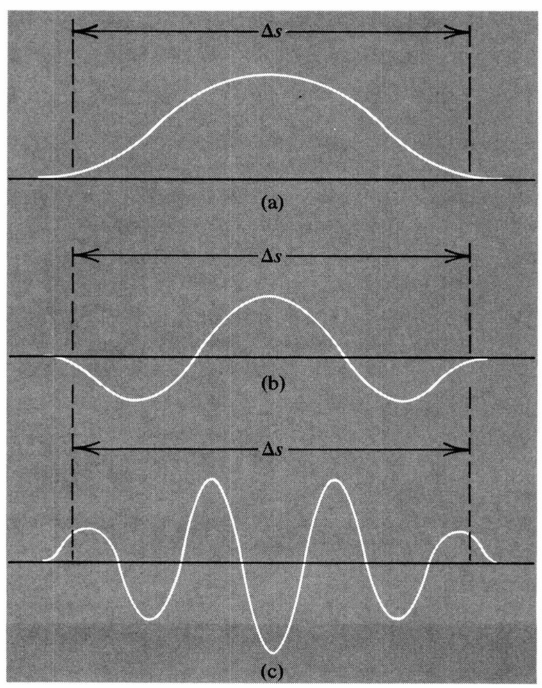


Figure 6·7 Oscillators in a wave train have opposite displacements if their separation is ½ wavelength.

Figure 6·8 Pulse patterns of disturbance in a medium. The approximate length of the pulse is denoted by Δs.



frequency (symbol f). The period and frequency are reciprocals of one another (Eq. 6.2), just as are the wavelength and wave number. The period and frequency describe the time variation of the oscillator displacements, while the wavelength and wave number describe the spatial variation.

Wave speed. One of the most striking properties of waves is that they give the appearance of motion along the medium. If you look at the pattern of displacements at two successive instants of time (Fig. 6.5), you see that the wave pattern appears to have moved to the right (along the medium), although the individual oscillators have only moved up and down. Since the pattern actually moves, you can measure its speed of propagation through the medium. The wave speed is usually represented by the symbol v (Section 2.2).

You can conduct a thought experiment with the oscillator model for the medium to find a relationship among period, wavelength, and wave speed. Imagine the oscillator at a wave crest carrying out a full cycle of its motion (Fig. 6.6). While this goes on, all the other oscillators also carry out a full cycle, and the wave pattern returns to its original shape. The wave crest that was identified with oscillator A in Fig. 6.6, however, is now identified with oscillator B. Hence the wave pattern has been displaced to the right by 1 wavelength. The wave speed is the ratio of the displacement divided by the time interval (Eq. 2.2), in this instance the ratio of the wavelength divided by the period ($\lambda/\mathcal{T}$, Eq. 6.3a). By using Eq. 6.2, $f = 1/\mathcal{T}$, you can obtain the most useful form of the relationship: $v = \lambda f$, or wave speed is equal to wavelength times frequency (Eq. 6.3b).

Positive and negative displacement. Waves are patterns of disturbances of oscillators from their equilibrium positions. The displacement is sometimes positive and sometimes negative. In Fig. 6.3, the open circles and the horizontal line drawn along the middle of the wave show the equilibrium state of the medium. Displacement upward may be considered positive, displacement downward negative. In water waves, for example, the crests are somewhat above the average or equilibrium level of the water and the troughs are somewhat below the average or equilibrium level of the water. In fact, the water that forms the crests has been displaced from the positions where troughs appear.

By definition, the pattern in a wave train repeats itself after a distance of 1 wavelength. It therefore also repeats after 2, 3, ... wavelengths. Consequently, the oscillator displacements at pairs of points separated by a whole number of wavelengths are equal. If you only look at a distance of 1/2 wavelength from an oscillator, however, you find an oscillator with a displacement equal in magnitude but opposite in direction (Fig. 6.7).

Wave pulses. In the *wave trains* we have been discussing, a long series of waves follow one another, and each one looks just like the preceding one. On the other hand, a ***wave pulse*** is also a disturbance in the medium but it is restricted to only a part of the medium at any one time (Fig. 6.8). It is not possible to define frequency or wavelength for a pulse since it does not repeat itself. The concept of wave speed,

Equation 6.2 (period and frequency of a wave)

period (time for one complete repetition, in seconds) $= \mathcal{T}$
frequency (number of complete repetitions in one second, per second) $= f$

$$\mathcal{T} \times f = 1,\ f = \frac{1}{\mathcal{T}},\ \mathcal{T} = \frac{1}{f}$$

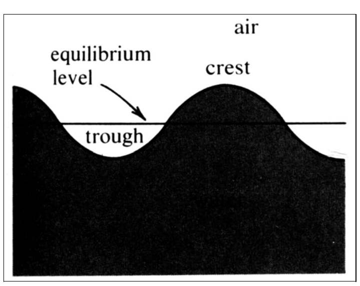
EXAMPLES

If $\mathcal{T} = 0.05$ sec,
$$f = \frac{1}{\mathcal{T}} = \frac{1}{0.05} = 20/\text{sec}.$$
If $\mathcal{T} = 3.0$ sec,
$$f = \frac{1}{\mathcal{T}} = \frac{1}{3.0} = 0.33/\text{sec}.$$
If $\mathcal{T} = 10^{-6}$ sec,
$$f = \frac{1}{\mathcal{T}} = \frac{1}{10^{-6}} = 10^{6}/\text{sec}.$$

Equation 6.3 (wave speed)

wave speed $= v$

$$v = \frac{\Delta s}{\Delta t} = \frac{\lambda}{\mathcal{T}} \qquad (a)$$
$$v = \lambda f \qquad (b)$$



however, is applicable to pulses since the pulse takes a certain amount of time to travel from one place to another. In Section 6.2 we will describe how wave trains and wave pulses can be related to one another.

Examples of wave phenomena. The oscillator model for a medium can be applied to systems in which small deviations from a uniform equilibrium arrangement can occur. One such system is a normally motionless water surface that has been disturbed so that water waves have been produced. Another example is air at atmospheric pressure in which deviations from equilibrium occur in the form of pressure variations: alternating higher or lower pressure. Such pressure variations are called sound waves. A third example is an elastic solid such as Jell-O, which can jiggle all over when tapped with a fork. In the oscillator model, movement results from oscillating displacements within the Jell-O after the fork displaced the oscillators at the surface.

Oscillator model for sound waves. Since sound in air is of special interest, we will describe an oscillator model for air in more detail. Visualize air as being made up of little cubes of gas (perhaps each one in an imaginary plastic bag). When acted upon by a sound source, the first cube is squeezed a little and the air inside attains a higher pressure (Fig. 6.9). The first cube then interacts with the next cube by pushing against it. After a while the second cube becomes compressed and the first one has expanded back to and beyond its original volume. The second cube then pushes on the third, and so on. In this way the sound propagates through the air.

The initial pressure increase above the equilibrium pressure may be

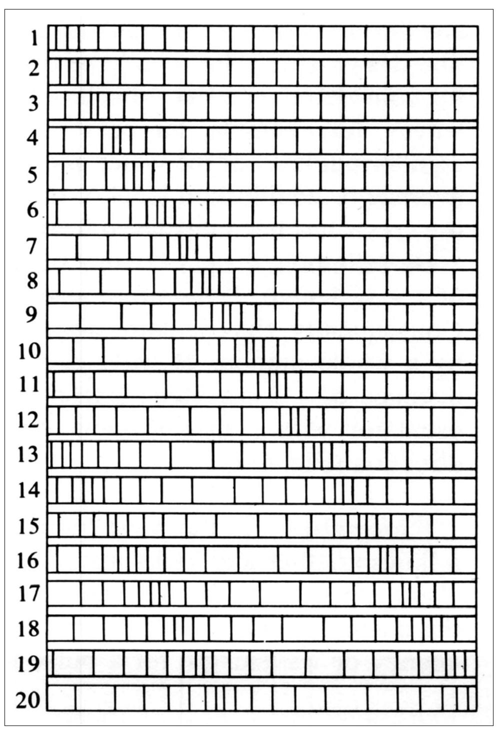


Figure 6.9 A gas bag model for air is used to represent the propagation of a sound wave. An individual bag of gas is alternately compressed and expanded. Its interaction with adjacent bags of gas leads to propagation of the compression and expansion waves.

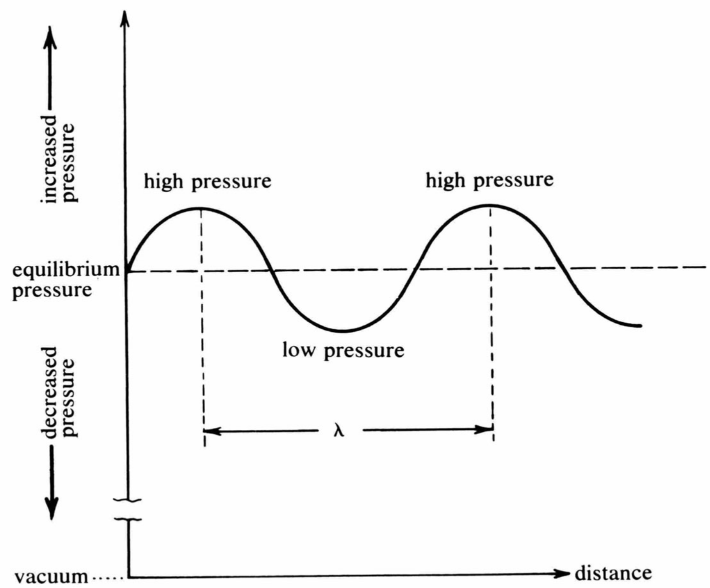


Figure 6.10 Pressure profile in a sound wave. The graph shows deviations from the equilibrium pressure.

created by a vibrating piano string or a vibrating drumhead. In addition to regions of increased pressure, the sound wave also has regions of deficient pressure where the air has expanded relative to its equilibrium state.

Thus the sound wave consists of alternating high-pressure (above equilibrium) and low-pressure (below equilibrium) regions. A pressure profile (pressure versus distance) for a pure tone has the typical wave pattern shown in Fig. 6.10.

6.2 Superposition and interference of waves

The superposition principle. Can you visualize what happens when two waves overlap? In the oscillator model, it is easy to describe the medium at a place where there are two or more waves at the same time. Each oscillator is displaced from its equilibrium position by an amount equal to the sum of the displacements associated with the waves separately (Fig. 6.11). In other words, you visualize the oscillator displacements associated with each of the wave patterns and add them together. This procedure takes for granted that the waves do not interact with one another, but that each propagates as though the others were not present.

The property of non-interaction we have just described is called the *superposition principle*. It makes the combination of waves simple to carry out in thought experiments, and it has been exceedingly valuable for this reason. Fortunately, a wave model that incorporates the superposition principle describes quite accurately many wave phenomena in nature.

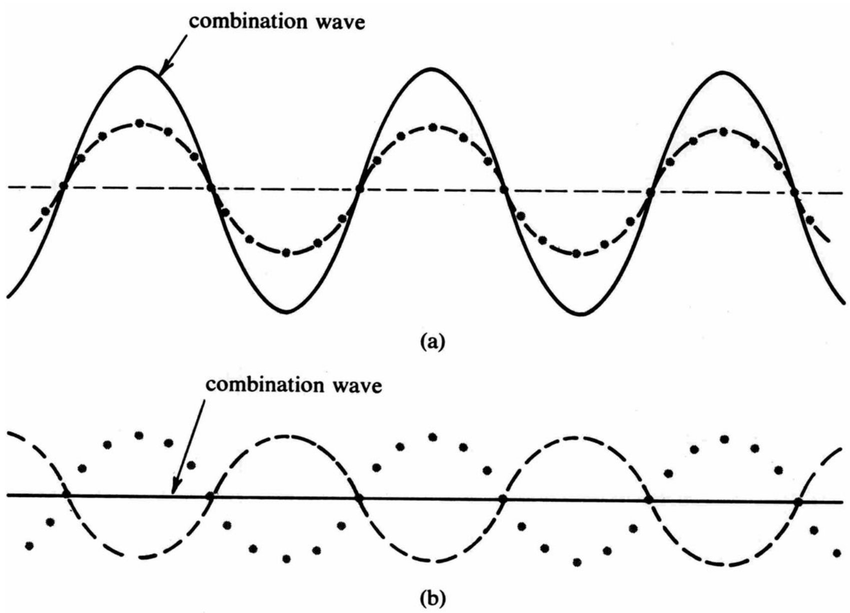


Figure 6.11 Superposition of two waves leads to interference. One wave is represented by black dashes, the other by dots. The combination wave is the sum of both waves and is represented by the solid line.

(a) Constructive interference occurs when dotted and dashed waves reinforce each other.

(b) Destructive interference occurs when dotted and dashed waves cancel each other.

Interference of waves. Consider now what may happen to the oscillator motion as a result of the superposition of two waves. The two waves may combine in various ways. Perhaps each of two wave patterns has an upward displacement of an oscillator at a certain time and at a certain place. In such a case, the upward displacement in the presence of the combined wave will be twice as big as that from one wave alone (Fig. 6.11(a)). If there are simultaneous downward displacements in the two waves separately, the combined displacement will be twice as far down. Suppose you consider a point in space where one wave has an upward displacement and the other wave has an equal downward displacement at the same time. Now, the upward (positive) displacement and the downward (negative) displacement add to give zero combined displacement (zero amplitude of oscillation). In fact, it is possible for

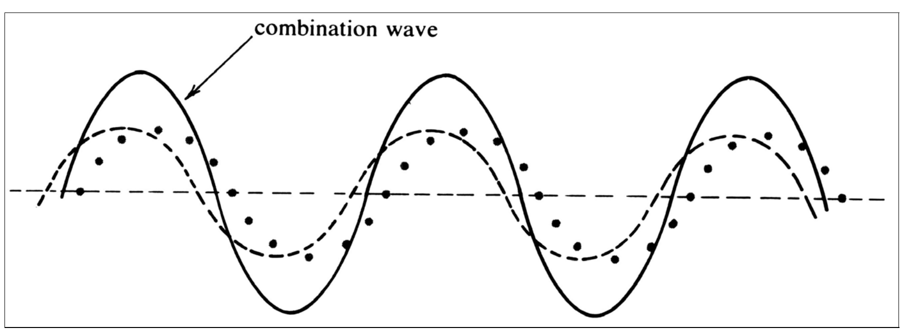
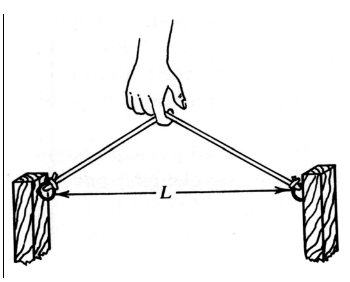


Figure 6.12 Superposition of two waves leading to partially destructive interference. The displacements of the dashed and dotted waves are added together at each point to yield the displacement of the combination wave (represented by the solid line). Note that displacement below the line is negative.

The "one-particle model" for a real object is a "very small object that is located at the center... of the region occupied by the real object." (Section 2.1) This is a way to think about an object so as to focus on the object's position, motion, and inertia without considering its shape and orientation. Complex objects can be thought of as two or more particles that interact in defined ways, or with the many-interacting-particles (MIP) model. In such models, each particle is thought of as a single, tiny bit of matter. The matter itself is thought of as indestructible, or "conserved." Such particles cannot "cancel" one another to cause destructive interference.

A wave, on the other hand, is quite different. A wave, as explained in this chapter, is thought of as a disturbance or oscillation that passes through matter. The displacement of the particles can be positive or negative and, as with (+1) + (–1) = 0, two waves can cancel one another.

In the 20th century, physicists found that matter in the micro-domain behaves in ways that conform with neither the particle nor the wave model. This led to the "wave-particle duality" and quantum mechanics. (Chapter 8).



two waves to combine in such a way that they completely cancel one another, as in Fig. 6.11b.

This characteristic of waves makes their behavior different from what we expect of material objects, particularly when we think of them as single particles (Section 2.1) or as made up of particles. If one particle and another particle are combined, you have two particles, and you cannot end up with zero particles. Two or more waves, however, may combine to form a wave with larger amplitude, a wave with zero amplitude, or a wave with an intermediate amplitude (Fig. 6.12).

This result of the superposition of waves is a phenomenon called *interference*. If waves combine to give a larger wave than either one alone, you have *constructive interference*. If waves tend to cancel each other, you have *destructive interference*. There is a continuum of possibilities between the extremes of complete constructive interference shown in Fig. 6.11(a) and complete destructive interference shown in Fig. 6.11(b). With particles, the concept of destructive interference is meaningless in that the presence of one particle can never "cancel" the presence of another.

Standing waves. When two equal-amplitude wave trains of the same frequency and wavelength travel through a medium in opposite directions, their interference creates an oscillating pattern that does not move through the medium (Fig. 6.13). Such an oscillating pattern is called a *standing wave*. The points in a standing wave pattern where there are no oscillations at all are called *nodes*. At a node, there is always complete destructive interference of the two wave trains; the displacements associated with the two waves at the nodes are always equal and opposite. Because the waves move in opposite directions at the same speed, each node remains at one point in space and does not move; this is the reason behind the choice of name: a *standing* wave does not move.

You can see in Fig. 6.13 that the distance between two nodes must be exactly ½ wavelength. This holds true not only for the illustration but also for *all* standing wave patterns. The reasoning is as follows. At any node, the two wave displacements must always be equal and opposite to produce complete destructive interference. At a distance of ½ wavelength, the displacement associated with each wave has exactly reversed (as illustrated in Fig. 6.7). Thus, the two displacements must again be equal and opposite and again produce a node.

An easy way to set up standing waves is to place a reflecting barrier in the path of a wave. The reflected wave interferes with the incident wave to produce standing waves. The nodes are easy to find because the oscillators remain stationary at a node. This offers a convenient way to determine the wavelength: measure the distance between nodes and multiply by 2.

Tuned systems. It is very fruitful to pursue the standing wave idea one step further. Suppose an elastic rope is tied to a fixed support at each end and the middle is set into motion by being pulled to the side and released (see drawing to left). How will the rope oscillate? To solve this problem, think of the pattern as being made up of wave trains in

Equation 6.4 (Possible number of half wavelengths that fit within L)

$$L = \frac{1}{2}\lambda$$

or

$$L = 2\times\left(\frac{1}{2}\lambda\right) = \frac{2}{2}\lambda$$

or

$$L = 3\times\left(\frac{1}{2}\lambda\right) = \frac{3}{2}\lambda$$

or

$$L = 4\times\left(\frac{1}{2}\lambda\right) = \frac{4}{2}\lambda$$

or

$$L = 5\times\left(\frac{1}{2}\lambda\right) = \frac{5}{2}\lambda$$

. . . and so on

Equation 6.5 (wavelengths permitted on a tuned system, from above)

$\lambda = \frac{2}{1}L$, or

$\lambda = \frac{2}{2}L = L$, or

$\lambda = \frac{2}{3}L$, or

$\lambda = \frac{2}{4}L = \frac{1}{2}L$, or

$\lambda = \frac{2}{5}L$

and so on . . .

Equation 6.6 (finding frequency for a given speed and wavelength)

$$f = \frac{v}{\lambda}$$

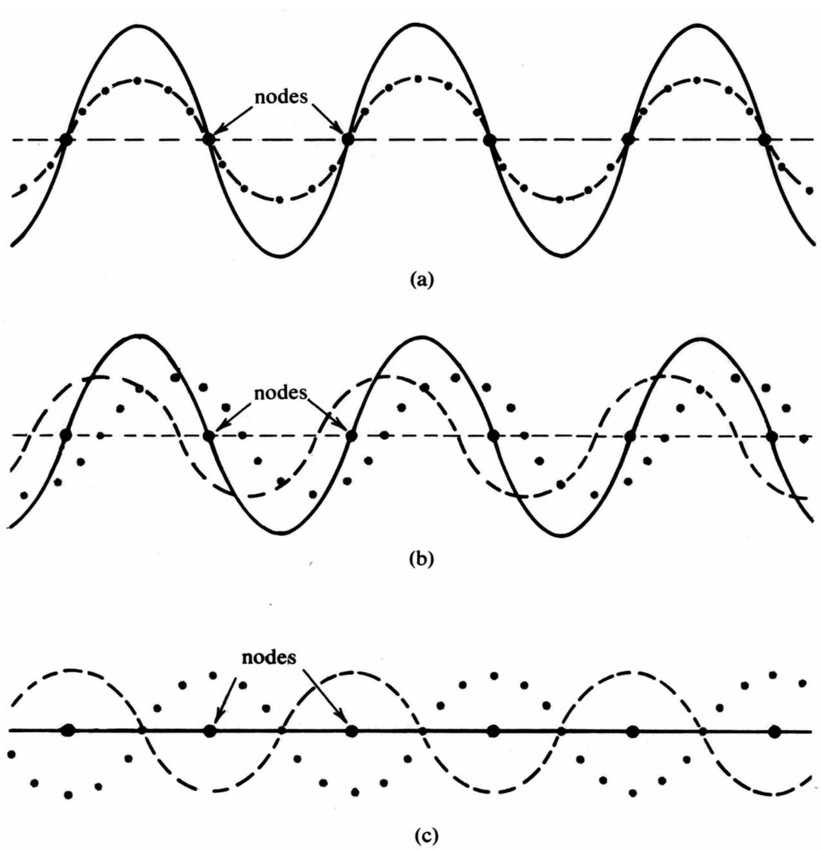


Figure 6.13 The formation of standing waves by the superposition of two wave trains propagating in opposite directions (dotted wave towards right and dashed wave toward left). The combination wave is the solid line. Note the stationary position of the nodes, marked by the large dots.
(a) Constructive interference of the two wave trains.
(b) Partially destructive interference after 1/8th of a period.
(c) Destructive interference after 2/8ths (1/4th) of a period.
Can you draw the pattern after 3/8ths of a period? After 4/8ths (1/2) of a period?

combinations, some moving to the right, others to the left. Because the ends are fixed, the wave pattern must be such that the ends of the rope are its nodes. The length of the rope is the distance between the nodes, which must be an integral multiple of ½ wavelength (Eq. 6.4). It follows that the wavelengths of the waves that can exist on this rope are related to the length of the rope by Eq. 6.5 to satisfy the conditions of nodes at the ends.

A system such as the rope with fixed ends is called a *tuned system*, because it can support only waves of certain wavelengths (Eq. 6.5) and the frequencies related to them by Eq. 6.6 (derived from Eq. 6.3b). The wave speed is a property of the medium from which the tuned system is constructed.

Musical instruments. Musical instruments employ one or more tuned systems whose frequencies are in a suitable relation to one another. For stringed instruments, such as the violin and guitar, the tuned system is a wire or elastic cord; for wind instruments, it is an air column in a pipe closed at one end; for drums, it is an elastic membrane whose edge is fixed; and so on.

The tone of the instrument is determined by the oscillation frequency of the tuned system. It is possible to change the frequency either through changing the length of the tuned system (and therefore changing the wavelength of the allowed standing waves) or through changing the wave velocity by modifying the medium in the tuned system.

Sound waves of a single frequency can be produced in closed pipes of a certain length. Longer pipes produce lower tones. A pressure wave starts at one end of the pipe and travels down the pipe, confined by the walls. When the wave reaches the other end of the pipe, it is reflected back and interferes with waves coming down the pipe. The interference forms a standing wave. This standing wave is of the characteristic wavelength determined by the length of the pipe and has the frequency that we hear.

Beats. Standing waves are created by the interference of waves with the same frequency. What will be the combined effect of two waves of differing frequencies? To answer this question, apply the superposition principle in a thought experiment in which two such waves are combined. Suppose the two waves are in constructive interference at one instant of time. Since one wave has shorter cycles than the other before repeating, they will soon get out of step. After a while, the two waves will be in destructive interference, and a little later in constructive interference again. So the net effect is an alternation from constructive interference (loud) to destructive interference (soft) and back again. These alternations in volume are called beats.

It is easily possible to calculate the time interval between two beats from the difference in frequency of the two interfering wave trains. During this time interval the two waves must go from constructive interference to destructive interference and back to constructive interference. Therefore, the higher-frequency wave must vibrate exactly once more than the lower frequency wave. The additional oscillation restores the constructive interference of the two waves, since waves repeat exactly after a whole oscillation. Hence the wave amplitude after the interval is equal to its value before, meaning that the next beat is ready to begin.

The number of oscillations made by either of the two waves is equal to its frequency (oscillations per second) times the time interval ($N_1 = f_1\Delta t$ and $N_2 = f_2\Delta t$, Eq. 6.7). The two numbers, according to the condition, must differ by one ($N_1 - N_2 = (f_1 - f_2)\Delta t$, Eq. 6.8). The conclusion is that the frequency difference times the time interval is equal to one ($\Delta f \, \Delta t = 1$, Eq. 6.9). The frequency of the individual waves determines the overall pitch of the sound, not the beat frequency; in fact, the beat frequency is $f_1 - f_2$.

Wave packets. Standing waves and beats are wave phenomena that are observable when two wave trains are combined. You may, of

Equation 6·7

frequencies of the two wave trains (per second) f_1, f_2
time interval (seconds) Δt
number of oscillations N_1, N_2

$$N_1 = f_1\Delta t, \qquad N_2 = f_2\Delta t$$

Equation 6·8

$$\begin{aligned} 1 &= N_1 - N_2 \\ &= f_1\Delta t - f_2\Delta t \\ &= (f_1 - f_2)\Delta t \end{aligned}$$

Equation 6·9

frequency difference Δf

$$\Delta f = f_1 - f_2 \quad \text{(a)}$$

$$\Delta f \Delta t = 1 \quad \text{(b)}$$

EXAMPLE

Frequencies of 255/sec and 257/sec

$$\Delta f = 2/sec$$

$$\Delta t = \frac{1}{\Delta f} = \frac{1}{2/sec} = 0.5\ sec$$

course, use the superposition principle and the rules for constructive and destructive interference to combine as many different wave patterns as you wish. In the early nineteenth century, it was discovered by Joseph Fourier (1768-1830) that any wave pattern could be formed by a superposition of one or more wave trains, as illustrated below. All wave phenomena can thereby be related to the frequencies, amplitudes, wavelengths, and velocities of the component wave trains in a wave pattern.

To illustrate Fourier's discovery, we will construct a wave pulse close to the one shown in Fig. 6.8a by combining the four wave trains drawn in Fig. 6.14. You are invited to read off the wave amplitudes from the graphs, to add the wave amplitudes of the four waves, and to verify that the combined wave drawn in Fig. 6.14 really is obtained by superposition of the four wave trains. By combining more and more wave trains of other wavelengths and successively smaller and smaller amplitudes, you can achieve further constructive and destructive interference at various locations in the pulse. In this way you could obtain a closer and closer approximation to the wave pulse shown in Fig. 6.8a and Fig. 6.14 (see Fig. 6.15).

The representation of wave pulses by a superposition of wave trains has led to the introduction of the suggestive phrase *wave packet* (instead of wave pulse), which we will also adopt. The superposition procedure can be quite tedious to work out in detail if many wave trains must be combined to achieve success. The essence of the procedure, however, is to select wave trains that interfere destructively in one wing of the wave packet, constructively at the center, and destructively again in the other wing. This can be achieved if one wave train has one more

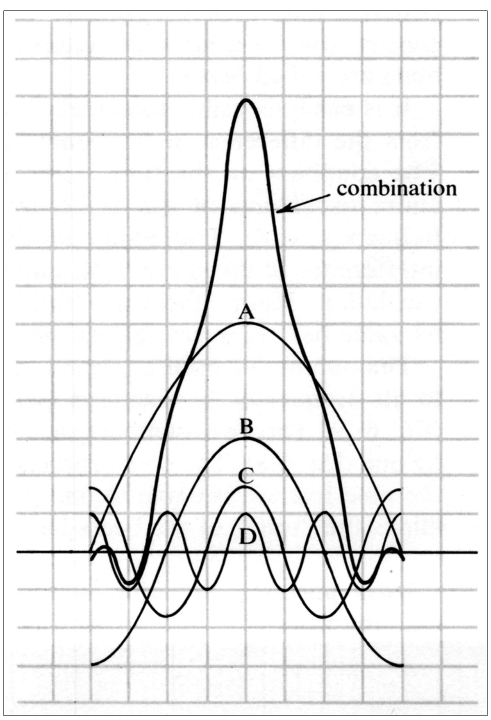
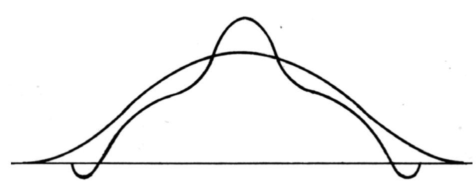
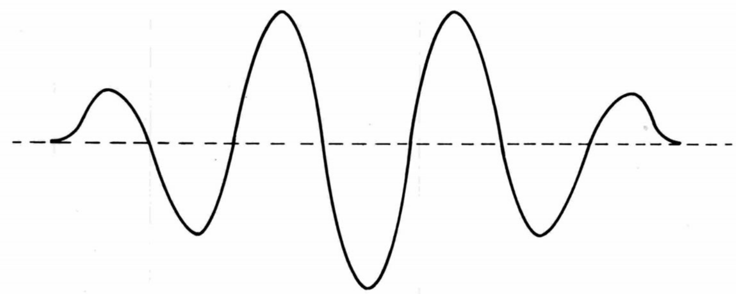


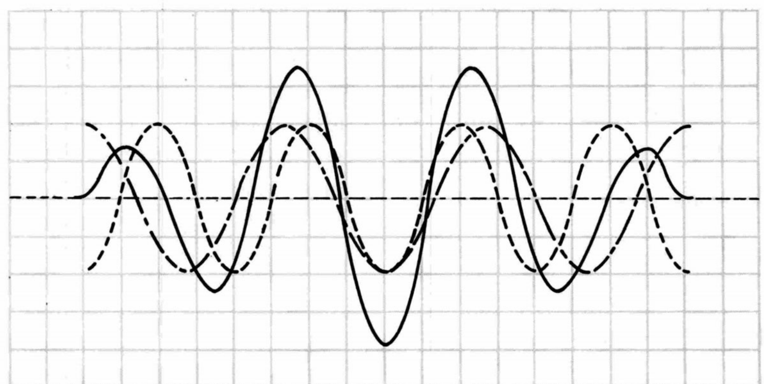
Figure 6.14 (left) The superposition of four wave trains to produce a wave pulse.

Figure 6.15 (below) The wave packet in Fig. 6.14 and the pulse in Fig. 6.8a have been drawn to the same scale for easier comparison.





(a)



(b)

Figure 6.16 The superposition of wave trains to produce a wave packet.
(a) The wave packet pictured in Fig. 68c, enlarged.
(b) A very similar wave packet constructed by the superposition of two wave trains.

full wave over the length of the packet than does the other one. Look, for example, at the wave packet with about four ripples shown in Fig. 6.8c, and reproduced here (Fig. 6.16a). We can combine two wave trains (one with four full waves over the length of the packet and one with three) to find a first approximation to the desired wave packet (Fig. 6.l6b).

Uncertainty principle. We will now formulate a general principle governing the superposition of wave trains to form wave packets. It is called the *uncertainty principle*, and it has played a very important role in the application of the wave model to atomic phenomena, which we will describe in Chapter 8.

Physical significance. The content of the uncertainty principle is that a wave packet that extends over a large distance in space (large Δs) is obtainable by superposition of wave trains covering a narrow range in wave numbers, but that a wave packet that extends over only a short distance in space (small Δs) must be represented by the superposition of wave trains covering a wide range in wave number. It is consequently impossible to construct a wave packet localized in space (small Δs) out of wave trains covering a narrow range in wave number. This idea is known as the "uncertainty principle" because it means that there is an inherent uncertainty in our ability to measure the exact position of a wave packet; Δs represents the "length" of the wave packet and thus the range of uncertainty in our measurement of the packet's position. The size of Δs is closely related to the range of wave numbers included in the packet. We cannot specify the range of wave numbers (Δk) precisely, but we can relate it to the size (Δs) of the wave packet. We will now derive a mathematical model that expresses this relationship.

Equation 6·10

wave number of the two wave trains (per meter) k_1, k_2
wave packet length (meters) Δs
number of waves N_1, N_2

$$N_1 = k_2\Delta s, \qquad N_2 = k_2\Delta s$$

Equation 6·11

$$\begin{aligned} 1 &= N_1 - N_2 \\ &= k_1\Delta s - k_2\Delta s \\ &= (k_1 - k_2)\Delta s \end{aligned}$$

Equation 6·12

wave number difference Δk

$$\Delta k = k_1 - k_2 \quad \text{(a)}$$

$$\Delta k\Delta s = 1 \quad \text{(b)}$$

Mathematical model. The calculation proceeds in the same way as the calculation for the time interval between beats in Eqs. 6.7, 6.8, and 6.9. First, we select two wave trains with different wave numbers k_1 and k_2, one a little larger and one a little smaller than the average wave number of all the waves needed. Each wave train has a certain number of waves ($N_1 = k_1\Delta s$ and $N_2 = k_2\Delta s$) within the length (Δs) of the wave packet (Eq. 6.10). By how much do these two numbers have to differ? They have to differ sufficiently so that the two wave trains are in destructive interference in the regions to the left and to the right of the wave packet's center, where they are in constructive interference. The distance between the two regions is approximately the spatial length Δs of the wave packet. Now, to achieve the desired destructive interference in both regions, the wave train with the shorter wavelength has to contain at least one more whole wave than the other in the distance Δs, that is: $1 + N_2 = N_1$, or $1 = N_1 - N_2$. This condition is applied in Eq. 6.11 to yield an important result: the range of wave numbers (Δk) times the width of the wave packet (Δs) is equal to one (Eq. 6.12b).

Comparison of beats and wave packets. It is clear that Eqs. 6.9 and 6.12b are closely similar. You may consider both of them as statements of an uncertainty principle for wave packets if you are willing to think of one beat pulsation as a wave packet. Equation 6.12 refers to the size of the wave packet in space. Equation 6.9 refers to the duration of the wave packet in time. The wave trains included in a wave packet have a certain average wave number or frequency, and extend above and below these average values by an amount equal to about one half of the wave number difference Δk or frequency difference Δf. The wave packet includes wave trains of substantial amplitude within this range of wave number or frequency, and wave trains of progressively smaller and smaller amplitude outside this range. The exact amplitude distribution of the included wave trains is determined by the shape of the wave packet and can be calculated by more complicated mathematical procedures developed by Fourier and later workers. We apply the uncertainty principle to wave packets below in Examples 6.1 and 6.2

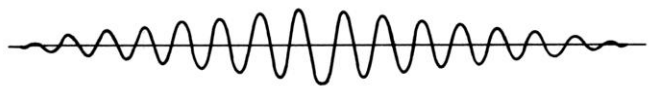
EXAMPLE 6.1. A telegraph buzzer operates at a pitch of 400 vibrations per second. A sound wave packet is formed by depressing the key for 0.1 second. What is the frequency range in the wave packet?

Solution:

$\Delta f\, \Delta t = 1$, $\Delta t = 0.1$ sec., hence $\Delta f = (1/\Delta t) = (1/0.1 \text{ sec}) = 10/\text{sec}$.

The frequency range is about 395 per second to 405 per second.

EXAMPLE 6.2. The wave packet pictured here is 0.08 meter long and contains approximately 16 ripples. What is the wave number range in this wave packet?



Solution : Average wave number k $= \dfrac{16}{0.008} = 200/m$

$$\Delta k\,\Delta s = 1,\ \Delta s = 0.08\ m,\ \Delta k = \frac{1}{\Delta s} = \frac{1}{0.08\,m} = 12/m$$

The wave number range is 194/m to 206 /m

6.3 Huygens' Principle

Ripple tank. Let us now return to study the propagation of waves by experimenting with water waves. A ripple tank is a useful device for observing water waves. It is a shallow tank with a glass bottom through which a strong light shines onto a screen (Fig. 6.17). Dipping a wire or paddle into the water, creates waves on the water surface; the crests of the waves create bright areas on the screen and troughs create shadows. The patterns of disturbance of the water surface may be observed (Fig. 6.18). A wide paddle generates straight waves (Fig. 6.18a), while the point of a wire generates expanding circular waves (Fig. 6.18b).

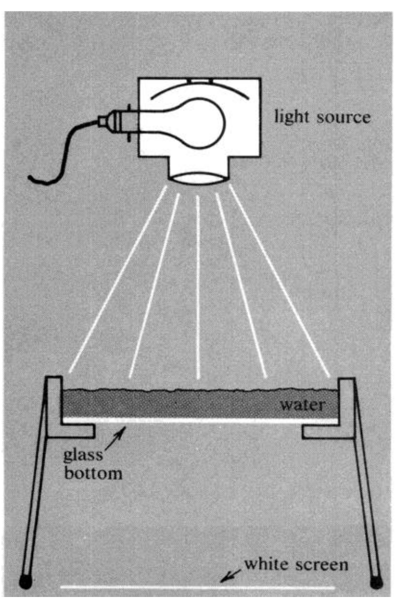


Figure 6.17 (left) Diagram of a ripple tank used for the production and observation of water waves. The wave crests and troughs create bright areas and shadows on the screen.

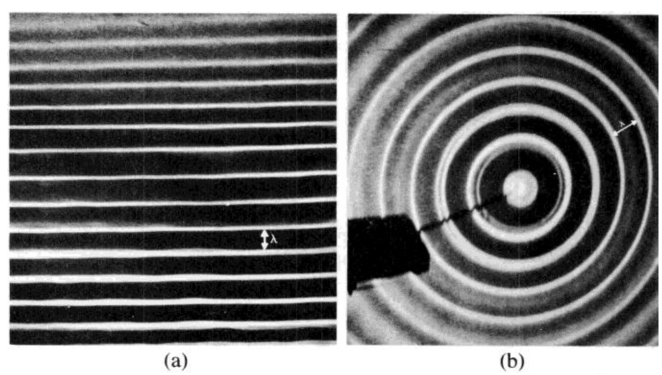


Figure 6.18 (above) Water waves in a ripple tank. (a) Waves generated by a wide paddle. (b) Waves generated by the point of a wire.

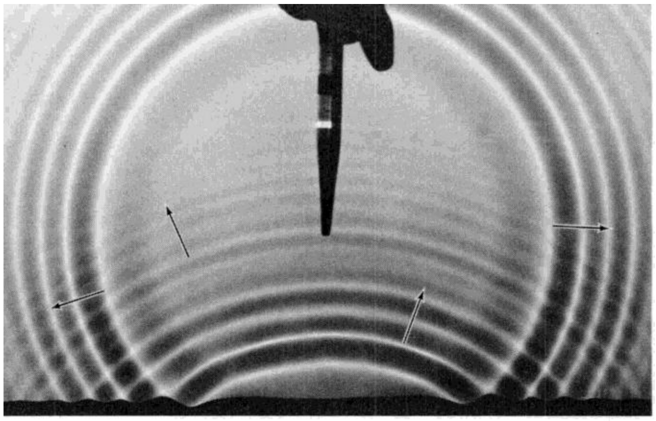


Figure 6.19 The bright lines of the wave crests indicate the wave fronts. The arrows at right angles to the wave fronts indicate the direction of propagation. The waves were originally produced by the tip of the pointer at the center of the photo. The wave fronts form circles centered on the point where they were created until they reflect from the barrier at the bottom of the photo. Where does the wave appear to be diverging from after it is reflected? Can you relate this to what you see in a plane (flat) mirror, as in Fig. 5.17?

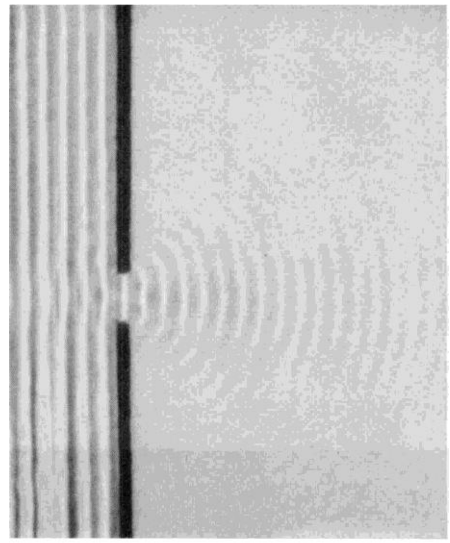


Figure 6.20 Straight waves from the left impinge on a barrier with a hole. Note the curved, circular shape of the wave front to the right of the barrier

The point of a wire may be considered as a point source of waves. The reflection of a circular wave pulse created by a pencil point touched to the water surface is shown in Fig. 6.19.

Wave patterns. To describe the pattern, we identify the *wave front*, which is the line made by each wave crest or trough, and the *propagation direction* in which the wave is traveling. The wave always travels in the direction at right angles to the wave front. Therefore, the wave travels in different directions at different parts of a curved wave front such as the one shown in Fig. 6.19.

You can make an interesting discovery if you use a barrier to block off all but one small section of the water. The waves passing through the hole from one side of the barrier to the other spread out in ever increasing circles (Fig. 6.20). This shows a very important result: the small section of the wave front acts as if it were itself a point source of waves.

Huygens' wavelets. In the oscillator model, the oscillator in the small section moves in rhythm with the waves impinging from the source side of the barrier; it also interacts with the oscillators on the other side and

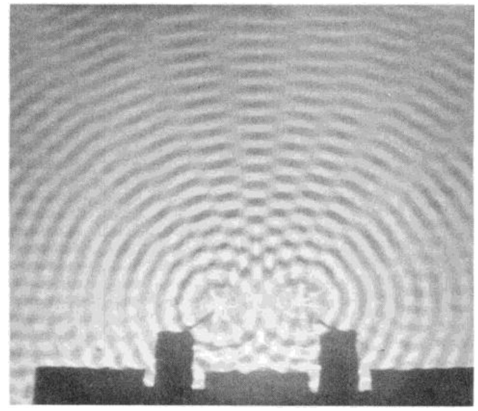


Figure 6.21 Two point sources produce an interference pattern. Note the lines of "nodes" fanning out from the sources.

sets them in motion as though it were a point source. In fact, you can think of every point of a wave front as the source of *wavelets* (numerous mini-waves generated by another wave) that radiate out in circles. That is, each oscillator interacts equally with the other oscillators in all directions from it. This principle is called *Huygens' Principle*. The wavelets have the same frequency of oscillation as their source points in the old wave front. When a wave front encounters a barrier, then most parts of the wave front are prevented from acting as wave sources. What remains is the circular wavelet originating from that part of the wave front that passes through the hole in the barrier.

Two-hole interference. When the barrier has two holes, the waves not only pass through both holes and spread out, but there also is interference between the waves coming from these two "sources." The observable result is very similar to the interference produced by waves from two adjacent point sources (Fig. 6.21). Note the lines of "nodes" fanning out at various angles from the sources, forming what is known as a "two-hole (or double-slit) interference pattern." This pattern demonstrates the existence of interference and can be observed in all waves (including light and sound), not just those in a ripple tank.

Construction of wave fronts. The position of the wave front at successive times may be found by seeking the region of constructive interference of the wavelets emanating from all the source points in a wave front. When there is no barrier, the complete circular wavelets originating from each point in the wave front are not seen because of destructive interference among them.

Schematic diagrams for the procedure of locating the constructive interference are drawn in Fig. 6.22. These diagrams show a wave crest at three successive instants. Huygens' Principle is applied to source points a in the initial wave crest AB to obtain the circles b, c, d. The destructive and constructive interference of all these wavelets results in a new wave crest at the position of the common tangent line CD of all the circles. After a second equal time interval, all the circles

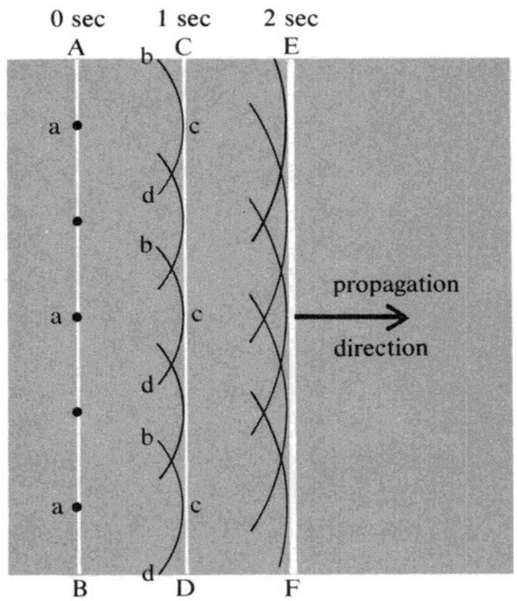


Figure 6.22 The wave front AB (thick white line) advances to CD and then EF, which are the common tangent lines of all the circular wavelets (thin black lines) from Huygens' sources (black dots) in the wave front at AB.

are twice as large, but again the interference effects result in a wave crest EF at the position of the common tangent line of all the larger circles. In this way the straight wave crest advances.

6.4 Diffraction of waves

It is clear from Fig. 6.20 that waves do not necessarily travel in straight lines. Even though the incident wave is headed to the right, the wave transmitted through the hole has parts that travel radially outward

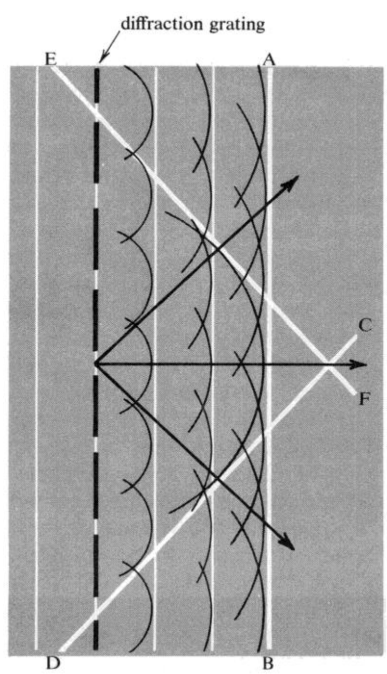


Figure 6.23 Huygens' Principle is used to find the waves transmitted by a diffraction grating. Note the wavelets (thin, curved black lines) centered on the slits. The white lines indicate the undiffracted wave crests (along common tangent line AB) and the diffracted wave crests (common tangent lines CD and EF). The black arrows show the directions of propagation of the observable waves.

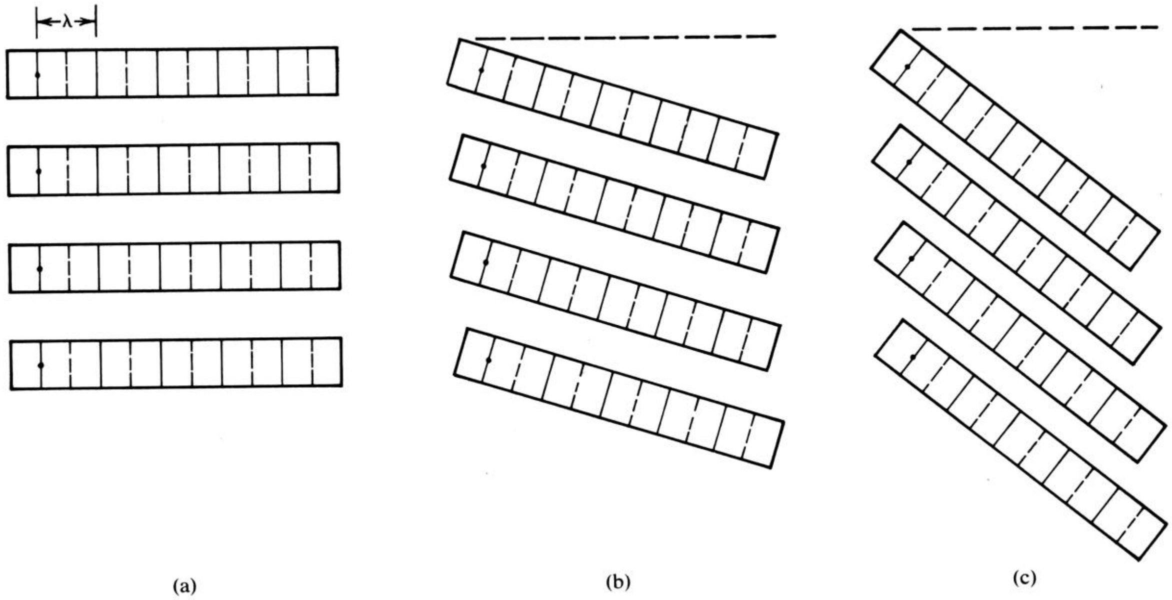


Figure 6.24 Paper strip analogue model for diffraction of waves by a grating. Four paper strips are marked at equal intervals to represent wave troughs and crests. The four strips are then pinned in a row to represent four wave trains passing through equidistant slits in a grating. The strips may be rotated, but are always kept parallel so that the strip direction represents the propagation direction. Interference is determined by the superposition of crests and troughs on the strips.
(a) Constructive interference in the un-diffracted direction is indicated by the alignment of crests with crests, troughs with troughs.
(b) Destructive interference is indicated by the alignment of the crests of one "wave" and the troughs of the adjacent "wave."
(c) Constructive interference in the diffracted direction is indicated by the alignment of crests with crests, troughs with troughs.

from the hole. In other words, the wave was deflected (or bent) by the barrier. This process of deflection of waves passing beside barriers is called *diffraction*. Diffraction makes it possible for waves to bend around a barrier.

Diffraction grating. Let us now apply Huygens' Principle to a device called a *diffraction grating*. A diffraction grating has many evenly spaced slits through which waves can travel. Between the slits, waves are absorbed or reflected. A wave coming through the slits radiates out from each slit in circular wavelets according to Huygens' Principle (Fig. 6.23). You do not observe simple circular waves, however, because the many waves interfere, sometimes constructively and sometimes destructively. A convenient analogue model for diffraction that can be constructed from four strips of paper is described in Fig. 6.24.

(a)

(b) [Enlarge-ment of (a)]

(c) [Enlarge-ment of (b)]

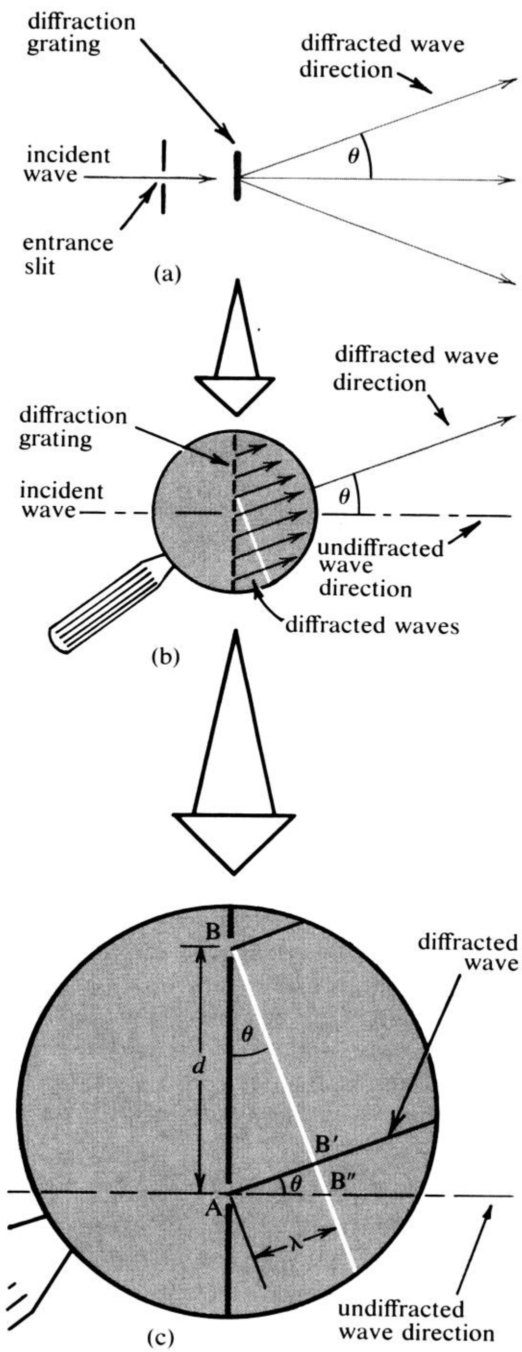


Figure 6·25 Construction of a mathematical model for diffraction by a large diffraction grating.
(a) Waves impinge on the grating from the left. Part of the wave pattern is diffracted at an angle θ, part continues in the undiffracted direction.
(b) Enlarged view of the grating shows that waves passing through adjacent slits travel different distances to contribute to the same wave front (white line).
(c) Constructive interference of diffracted waves occurs if the wave trains from adjacent slits are exactly 1 wavelength out of step as they contribute to one wave front (white line).

distance between slits	d
wavelength	λ
diffraction angle for constructive interference	θ
additional path length for waves passing slit A compared to waves passing slit B	$\overline{AB'}$

Diffraction condition: $\overline{AB'} = \lambda$

Step I: By definition, sine $\angle ABB' = \overline{AB'}/\overline{AB} = \lambda/d$.
Step II: Prove $\angle ABB' = \theta$.
(i) Extend line BB′ to the undiffracted wave direction at B″.
(ii) θ *is complementary to* $\angle AB''B'$ *in right triangle AB″B′.*
(iii) $\angle ABB'$ *is complementary to* $\angle AB''B'$ *in right triangle AB″B.*
(iv) Hence $\angle ABB' = \theta$.
Step III: It follows from I and II that sine $\theta = \lambda/d$.

With a large diffraction grating of many slits (perhaps 10,000 slits or more), constructive interference of the waves from all the slits occurs only when the adjacent strips are exactly one, two, or three waves out of step. For all other directions, you can find pairs of close or distant slits that give complete destructive interference and thereby cancel one another's wavelets. Waves are therefore diffracted by the grating only

Equation 6.13
(diffraction grating)

distance between slits (meters) $= d$
diffraction angle $= \theta$

$$\text{sine } \theta = \frac{\lambda}{d}$$

into certain special directions. The diffraction angle can be calculated from the condition for constructive interference (Fig. 6.25).

The diffraction grating formula states that the sine of the angle of diffraction is equal to the ratio of the wavelength to the distance between slits. (See Eq. 6.13 and Example 6.3.) The most important practical application of the diffraction grating has been to the study of light, which will be described in the next chapter.

EXAMPLE 6·3. Use of the diffraction formula.

(a) $\lambda = 0.2$ m, $d = 0.3$ m, $\theta = ?$

$$\text{sine } \theta = \frac{\lambda}{d} = \frac{0.2\text{m}}{0.3\text{m}} = 0.67$$
$$\theta = 42°$$

(b) $d = 10^{-6}$ m, $\theta = 25°$, $\lambda = ?$

$$\text{sine } 25° = 0.42$$
$$\lambda = d \text{ sine } \theta = 10^{-6} \text{ m} \times 0.42 = 0.42 \times 10^{-6} \text{ m}$$

(c) $\lambda = 10^3$ m, $\theta = 15°$, $d = ?$

$$\text{sine } \theta = 0.26$$
$$d = \frac{\lambda}{\text{sine } \theta} = \frac{10^3\text{m}}{0.26} = 3.9 \times 10^2 \text{ m}$$

(d) $\lambda = 10^{-4}$ m, $d = 10^{-2}$ m, $\theta = ?$

$$\text{sine } \theta = \frac{\lambda}{d} = \frac{10^{-4}}{10^{-2}} = 10^{-2}$$
$$\theta = 0.6°$$

(e) $\lambda = 0.3$ m, $d = 0.2$ m, $\theta = ?$

$$\text{sine } \theta = \frac{\lambda}{d} = \frac{0.3}{0.2} = 1.5$$

θ does not exist.

Diffraction by single slits and small obstacles. Huygens' Principle can also be applied to diffraction by a single slit opening (Fig. 6.20) and to diffraction by a short barrier. The result of the theory suggests that the ratio of the wavelength to a geometrical dimension of the diffracting barrier is of decisive importance for diffraction. In fact, if this ratio is very small (short wavelength, large slit, or large obstacle), the angles of diffraction are very small, so that diffraction is hardly noticeable. If the ratio is large (long wavelength, small slit, or small obstacle),

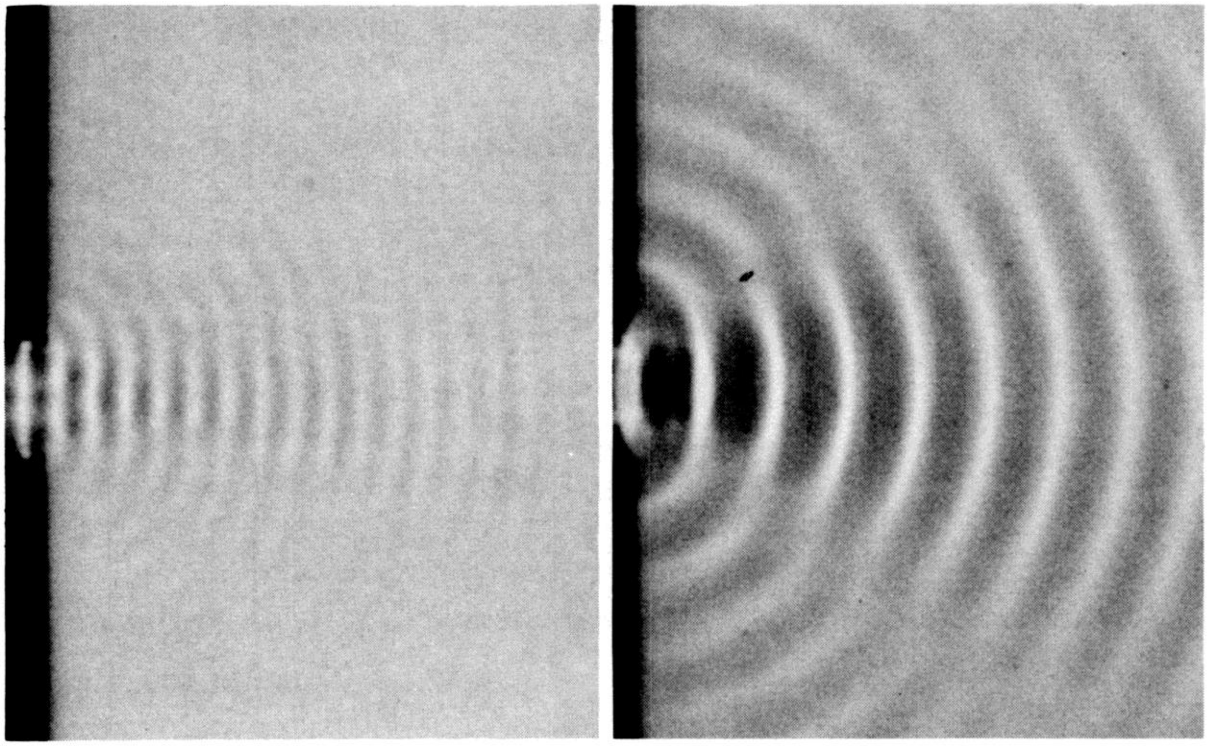
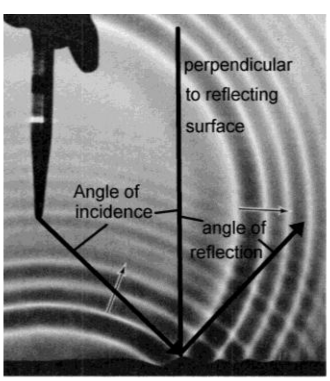


Figure 6.26 Diffraction of waves by an opening. In both photos, the waves are moving from left to right. In the left photo, the wavelength is relatively short (1/3 the width of the opening), so there is little diffraction, and only very weak waves are diffracted away from the original direction of propagation. In the right photo, the wavelength is longer (2/3 the width of the opening), and the waves experience substantial diffraction, spreading out in all directions. After passing through the opening, the wave fronts are essentially semicircles, showing that the waves are now moving in various directions away from the opening. This is a practical demonstration of Huygen's Principle: the waves passing through the opening act as point sources of new waves, which then travel in all directions away from the sources.

then diffraction covers all angles, but the amplitudes of the diffracted waves are very small because the slit or obstacles are small. For intermediate values of the ratio (wavelength comparable to the slit or obstacle in size), diffraction is an important and easily noticeable phenomenon. Two photographs of waves in a ripple tank (Fig. 6.26) show long and short wavelength waves passing through an opening and being diffracted when they pass through an opening. The greater diffraction of the longer wavelength waves is obvious.



6.5 Reflection of waves

The ripple tank photograph to the left (from Fig. 6.19) shows reflection of an expanding circular wave packet. We picked one point on the barrier and drew arrows showing the approximate direction of propagation before and after reflection from that point. The angles of incidence and reflection (as defined in Fig. 5.11) are shown. You can measure the angles to test whether they are equal; we measured one to be 43.5° and the other to be 45°; this is satisfactory agreement given the accuracy of our measurements.

We can also use Huygens' Principle to investigate the relation of these angles in a more general way. According to this principle, each point in a wave front acts like a source of wavelets propagating outward. The wavelets have the same frequency and wavelength as the original waves. The common tangent line of the wavelets is the wave front they produce by constructive interference.

The reflection process is illustrated in Fig. 6.27. A straight wave is incident on the reflecting barrier obliquely from the left. Between the

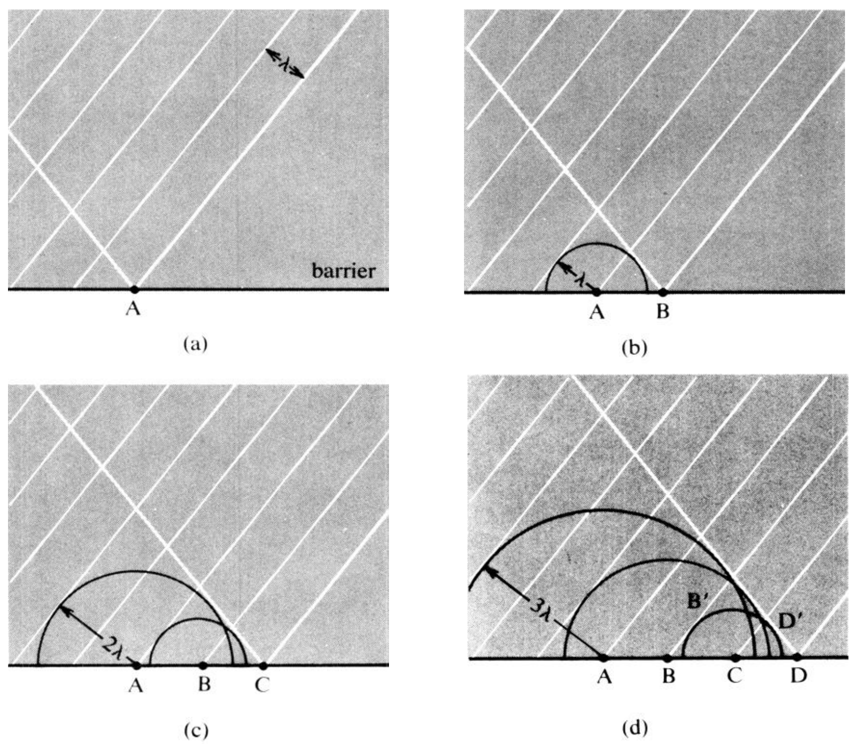


Figure 6.27 Reflection of waves by a barrier. The incident wave crests (white lines) are advancing toward the lower right. Only one reflected wave, moving toward the upper right, is shown. To avoid clutter in the diagram, we have not drawn the other reflected waves.

(a) Three wave crests are striking the barrier, which has reflected a section of the first crest. Point A on the first crest acts as a source of Huygens' wavelets.

(b) The wave crests advance by a distance of one wavelength (λ), and the wavelet from Point A has expanded into a semicircle of radius λ. Point B becomes a source of wavelets.

(c) The wave crests advance by another wavelength; the wavelet from Point A now has a radius of 2λ; the wavelet from B has a radius λ. and the Point C becomes a source of wavelets.

(d) The wavelet from A has radius 3λ; the wavelet from B has radius 2λ, the wavelet from C has radius λ, and Point D becomes a source of wavelets.

The wavelets constructively interfere all along the common tangent line DD', which defines the location and direction of the reflected wave fronts.

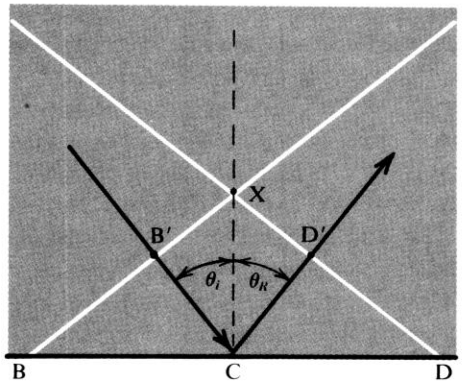


Figure 6·28 Construction of a mathematical model for wave reflection based on Fig. 6·27. Arrows B′C and CD′ represent, respectively, the incident and reflected propagation directions. They are at right angles to the corresponding wave fronts BB′ and DD′ (white lines). Consider the two right triangles XCB′ and XCD′. They share the common hypotenuse XC and have two sides equal, CB′ = CD′ = λ. Hence the two triangles are congruent. It follows that corresponding angles are equal, $\theta_i = \theta_R$.

four successive instants shown in Fig. 6.27, the wave advances between each drawing by 1 wavelength. The Huygens wavelets formed by the first wave crest passing through points A, B, C, and intermediate points on the barrier have a common tangent, which is the reflected wave front. The crests of the Huygens' wavelets all fall on the common tangent, where they interfere constructively; at all other points, the wavelets interfere destructively and cancel one other.

Equation 6.14 (Law of Reflection)

angle of incidence = θ_i

angle of refraction = θ_R

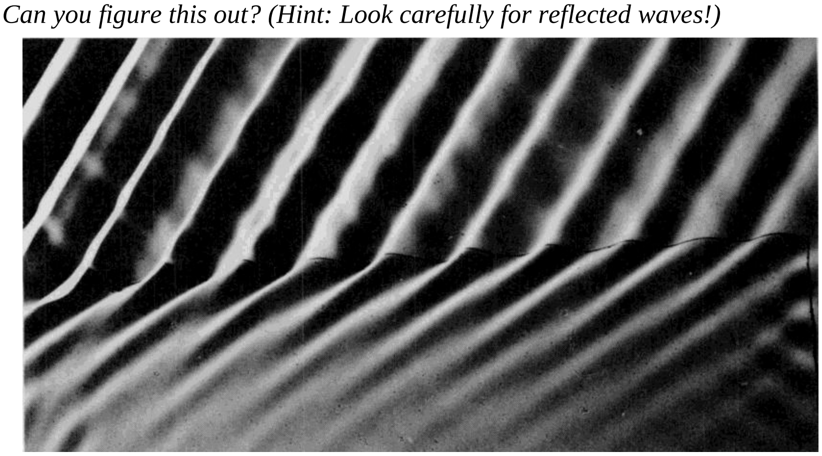
$$\theta_i = \theta_R$$

To relate the angles of incidence and reflection, the directions of propagation have to be taken into account. This is done in Fig. 6.28, where only one incident and one reflected wave crest from Fig. 6.27d are included. The application of Huygens' Principle in Fig. 6.28 results in a familiar conclusion: the angle of incidence is equal to the angle of reflection (Eq. 6.14). This statement may be called the *law of wave reflection*.

6.6 Refraction of waves

When a wave propagates from one medium into another, its direction of propagation may be changed. An example of this happening with water waves is shown in Fig. 6.29. The boundary here is between deep water above and shallow water below. Even though water is the

Fig 6.29 Water waves passing from a deeper region to a shallower region are refracted and travel in a different direction at the boundary. Huygens' Principle does not reveal which direction the waves are traveling. Can you figure this out? (Hint: Look carefully for reflected waves!)



Equation 6.3b

$v = f\lambda$

material on both sides of the boundary, it acts as a different medium for wave propagation when it has different depths. You can see that the wavelength is shorter in the shallow water and can infer from this that the wave speed is slower there (Eq. 6.3b).

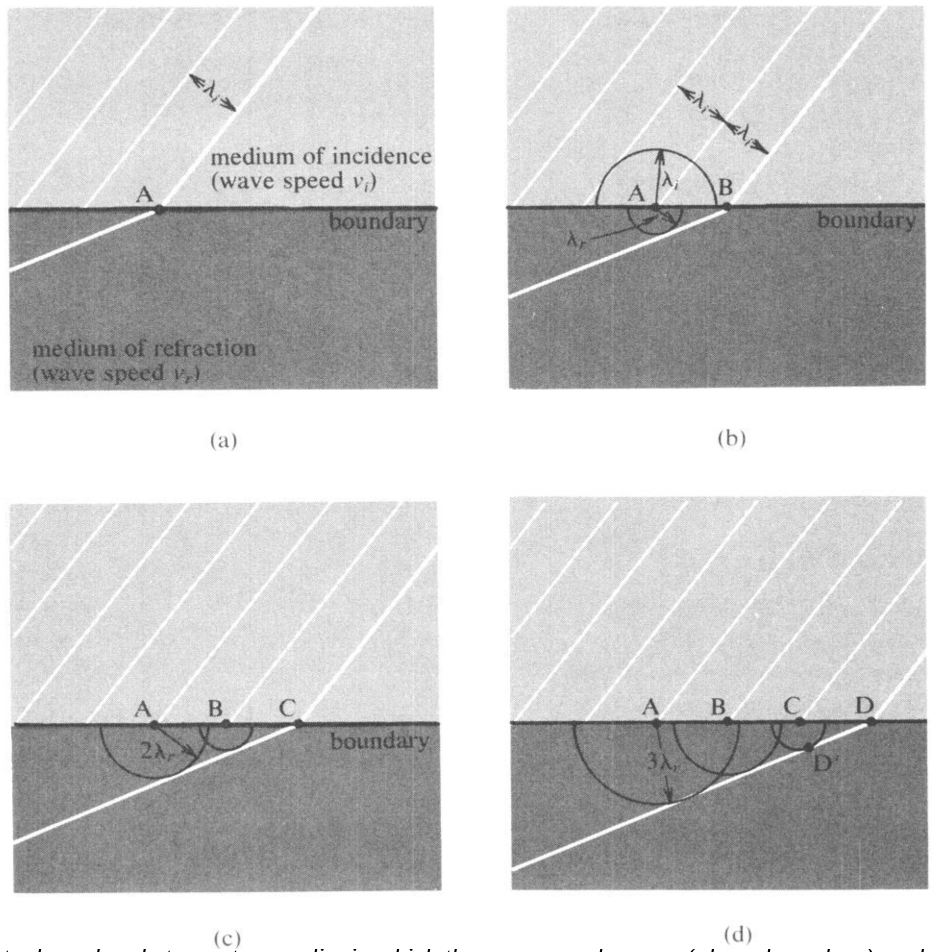


Figure 6.30 Refraction at a boundary between two media, in which the wave speeds are v_i *(above boundary) and* v_r *(below boundary). The wave speed is assumed to be* ***less*** *below the boundary than above it. (*v_r *is* ***less*** *than* v_i*). The waves are moving downward to the right.*

(a) Three wave crests (white lines) are incident on the boundary, which has refracted a section of the waves. Point A on the first crest acts as a source of Huygens' wavelets.

(b) The wavelet from point A has expanded into a semicircle of radius λ_r *below the boundary and a semicircle of radius* λ_i *above the boundary. The latter gives rise to a reflected wave (see Fig. 6.27) and will not be described further. The wave crests advance by the distance* λ_i*. Point B becomes a source of wavelets.*

(c) The wavelet from A has reached radius $2\lambda_r$ *, the wavelet from B has radius* λ_r *. Point C becomes a source of wavelets.*

(d) The wavelet from A has radius $3\lambda_r$ *, the wavelet from B has radius* $2\lambda_r$ *, and the wavelet from C has radius* λ_r *. The common tangent line DD' coincides with the refracted wave front. Note the change in the direction of propagation (which is perpendicular to the wave front): we can see from the diagram that the refracted wave (below the boundary) is traveling* ***slower*** *and in a direction* ***farther*** *from the boundary surface than the incident wave (that is,* ***closer*** *to the perpendicular to the boundary).*

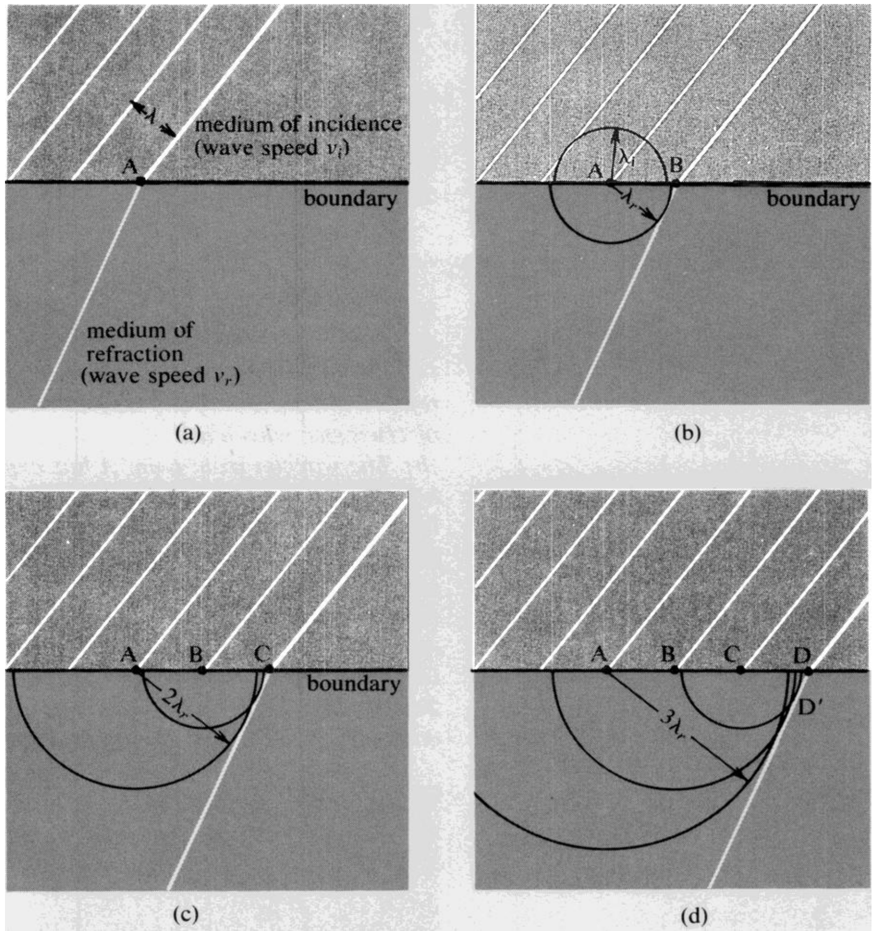


Figure 6.31 Refraction at a boundary between two media, in which the wave speeds are v_i (above boundary) and v_r (below boundary). The wave speed is assumed to be ***greater*** *below the boundary (v_r is* ***greater*** *than v_i). (This is similar to Figure 6.30 but the speeds are reversed.) The waves are assumed to be moving downward to the right.*

(a) Four wave crests (white lines) are incident on the boundary, which has refracted a section of the waves. Point A on the first crest acts as a source of Huygens' wavelets.

(b) The wavelet from point A has expanded into a semicircle of radius λ_r below the boundary and a semicircle of radius λ_i above the boundary. The latter gives rise to a reflected wave (see Fig. 6.27) and will not be described further. The wave crests advance by the distance λ_i. Point B becomes a source of wavelets.

(c) The wavelet from A has radius $2\lambda_r$; the wavelet from B has radius λ_r. Point C becomes a source of wavelets.

(d) The wavelet from A has radius $3\lambda_r$, the wavelet from B has radius $2\lambda_r$, and the wavelet from C has radius λ_r. The common tangent line DD' coincides with the refracted wave front. Note the change in the direction of propagation (which is perpendicular to the wave front): we can see from the diagram that the refracted wave (below the boundary) is traveling ***faster*** *and in a direction that is* ***closer*** *to the boundary surface than the incident wave (that is,* ***further*** *from the perpendicular to the boundary).*

Note: Although we have assumed above that the waves are traveling toward the right, this demonstration can also be carried out using the same diagram with the waves traveling in the opposite direction. Thus wave theory based on Huygens' Principle predicts that refracted waves will follow the same path in either direction. Does this seem reasonable to you? Can you suggest any observations or experiments that would confirm or refute this?

The refracting boundary. The change in the direction of propagation is called refraction, the same term that was introduced in Section 5.2. We will now find the law of refraction of waves by applying Huygens' Principle to the propagation of the wave across the boundary between two media with different wave velocities. Each point in the wave front that touches the medium of refraction acts like a source of wavelets that propagate into that medium. These waves have the same

Equation 6·15

wave speed in medium of incidence v_i
wave speed in medium of refraction v_r
wavelength in medium of incidence λ_i
wavelength in medium of refraction λ_r

$$v_i = \lambda_i f$$

$$v_r = \lambda_r f$$

Equation 6.17

$$\frac{\text{sine}\,\theta_\text{i}}{\text{sine}\,\theta_\text{r}} = \frac{\lambda_i}{\lambda_r} \quad (a)$$

$$\frac{\text{sine}\,\theta_\text{i}}{\text{sine}\,\theta_\text{r}} = \frac{v_i}{v_r} \quad (b)$$

frequency as their source, and therefore the same frequency as the wave in the medium of incidence. The wave in the medium of refraction, however, where the speed is different, has an altered wavelength, because wavelength, frequency, and speed are related by $v = \lambda f$ (Eq. 6.15). The ratio of the wavelengths in the two media is equal to the ratio of the wave speeds, since these two properties of the wave are directly proportional as long as the frequency remains the same (Eq. 6.16). Thus, the change in medium results in a changed wavelength.

Construction of the refracted wave front. The procedure for finding the law of refraction is very similar to that used in the preceding section to find the law of reflection. A straight wave is incident on the refracting boundary obliquely from the left. Between each of the four successive instants shown in Figs. 6.30 and 6.31, the wave advances by 1 wavelength. The Huygens' sources on the boundary generate wavelets that propagate into the second medium with the wave speed and therefore the wavelength appropriate to that medium. The case of reduced wave speed and wavelength is illustrated in Fig. 6.30, while the case of increased wave speed and wavelength is illustrated in Fig. 6.31. In both cases the wavelets originating in points A, B, and C (and intermediate points on the boundary) have a common tangent that is the refracted wave front.

Law of refraction. To relate the angles of incidence and refraction, the directions of propagation have to be taken into account. This is done for both cases above in Fig. 6.32, where only one incident and one refracted wave crest from the previous figures are included. The conclusion from the application of Huygens' Principle is that the sines of the angles of incidence and refraction have the same ratio as the wavelengths (Eq. 6.17a) and, therefore, the same ratio as the wave speeds in the two media (Eq. 6.17b). This result is similar in form to Snell's Law of Refraction: (n_i sine θ_i = n_r sine θ_r , Eq 5.2, Section 5.2), a key assumption in Newton's ray model of light. We shall study this further below in Section 7.2, where we will compare and evaluate the ray and wave models in some detail.

If you look at the propagation direction of the refracted waves in Figs. 6.30 and 6.31, you will recognize that the effect of crossing the boundary can be described as follows. In the medium with the slower wave, the propagation direction is farther away from the boundary surface; in the medium with the faster wave, the propagation direction is closer to the boundary surface. You may use the tables of the sine functions (Appendix, Table A.7) to solve problems on the refraction of waves.

Reflection at the boundary. The application of Huygens' Principle to the boundary between the two media leads to reflected wavelets as well as refracted ones. One reflected wavelet is indicated in Fig. 6.30b and one is indicated in Fig. 6.31b. Since these wavelets are in the medium of incidence, their speed and wavelength are appropriate to that medium. By pursuing their formation further, we could have obtained the same sequence of diagrams as are shown in Fig. 6.27. The wavelets would interfere constructively to form a reflected wave according to the law of reflection (Eq. 6.14). Thus wave theory suggests that we should also look for reflected waves, and, in fact, by looking carefully, you can indeed identify reflected waves in the deeper water of Fig. 6.29! In other

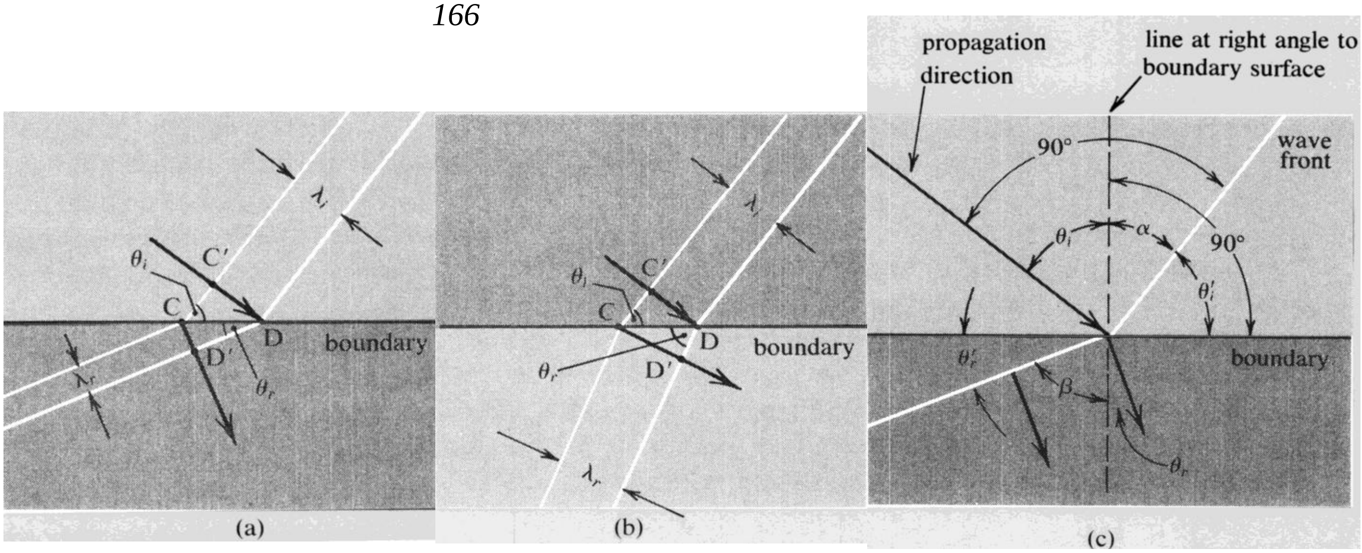


Figure 6.32 Construction of a mathematical model for wave refraction, based on Figs. 6.30 and 6.31. Note that in (a) the medium with the longer wavelength (faster speed) is at top in lighter shading. However, in (b) the medium with the longer wavelength (faster speed) is at bottom, also in lighter shading.
In both (a) and (b), Arrows C'D and CD' represent, respectively, the incident and refracted propagation directions. They are at right angles to the corresponding wave fronts (white lines) CC' and DD'. ASSUMPTION: Angle C'CD is equal to the angle of incidence θ_i and D'DC is equal to the angle of refraction θ_r ; we will prove this assumption in (c) below. The definition of the sine functions can be applied to right triangles CDC' and CDD' with the following results:

$$\text{sine } \theta_{\text{i}} = \frac{C'D}{CD} = \frac{\lambda_{\text{i}}}{CD} \quad (1)$$

$$\text{sine } \theta_{\text{r}} = \frac{D'C}{CD} = \frac{\lambda_{\text{r}}}{CD} \quad (2)$$

Divide Eq. (1) by Eq. (2)

$$\frac{\text{sine } \theta_{\text{i}}}{\text{sine } \theta_r} = \frac{\lambda_{\text{i}}}{\lambda_{\text{r}}} \qquad (3)$$

(c) Proof of ASSUMPTION asserted above: the angle of incidence θ_i equals the angle between the boundary and the wave front θ_i'. Two overlapping right angles in the medium of incidence are indicated in the figure (c) above. Both angles θ_i and θ_i' are complementary to the angle α. Consequently the two angles are equal, $\theta_i = \theta_i'$. The same construction with respect to the angle β in the second medium leads to the conclusion that $\theta_r = \theta_r'$.

words, the incident wave appears to be split by the boundary into a reflected wave and a refracted wave. Huygens' Principle has indeed served us well; however, it does not reveal how the energy carried by a wave is divided between reflection and refraction.

The occurrence of partial reflection gives an important clue about the sense of the direction of propagation of waves. By examining only the incident and refracted waves, such as in Fig. 6.29, you would not be able to determine whether the waves were incident as described at the beginning of this section (from upper left), or whether the waves were incident from the lower right and passed from the shallow water to the deeper water. The observation of reflected waves in the upper part of the photograph is evidence that the waves were incident from above.

Summary

The concept of waves has its roots in water waves. More generally, waves are oscillatory displacements of a medium from its equilibrium state. Two important forms that such disturbances can take are the wave train, in which the displacement pattern repeats over and over, and the wave pulse, in which the displacements are localized in space and time. The wavelength, wave number, period, frequency, amplitude, and speed of the waves can be defined for a wave train, but only the last two of these can be defined for a pulse. The frequency, wavelength, and speed of a wave train are related by v = λf (Eq. 6.3b).

Equation 6.3b (wave speed)

$$v = \lambda f$$

The wave theory is built upon the above ideas and applies to a wide variety of types of waves. The goal of wave theory is the construction of mathematical models to describe the behavior and propagation of waves. Wave theory explains and clarifies a large variety of phenomena. Such phenomena include sound, music, water waves, radio, light, constitution of the atom, traffic flow, and earthquakes.

The wave theory rests on two key assumptions about waves: 1) the superposition principle and 2) Huygen's Principle. The theoretical deductions from these assumptions can be compared with observation to identify the successes and the limitations of the wave theory.

According to the superposition principle, the displacement of the combination of two or more waves passing through the same point in space at the same time is the sum of the displacements of the separate waves. The result is constructive or destructive interference, depending on whether the separate waves reinforce or oppose one another.

Huygens' principle is used to investigate the propagation of waves. Each point in a wave front is considered as a source of circular outgoing wavelets. The amplitude and frequency of the wavelets are determined by the amplitude and frequency of the wave at the source point. The wavelets interfere constructively along their common tangent line, which is therefore the front of the propagating wave. Elsewhere, the wavelets interfere destructively and are not separately observable.

Huygens' Principle allows us to conduct thought experiments on the propagation of waves and furnishes a procedure for determining the results. We have used Huygens' Principle to understand diffraction, reflection, and refraction of waves.

The wave theory does not attempt to relate the wave speed, amplitude, and energy to properties of the medium, the wave source, and the wave absorber. These matters require more detailed working models for the three systems; their treatment is beyond the scope of this text.

List of new terms

medium (for wave propagation)
wave train
wave pulse
amplitude
frequency (f)
wavelength (λ)
wave number (k)
period ($\mathcal{T}$)
wave speed (v)
superposition
interference
constructive interference
destructive interference
node
tuned system
beats
wave packet
uncertainty principle
Huygens' Principle
wave front
propagation direction
Huygens' wavelets
diffraction
diffraction grating
reflection of waves
refraction of waves
standing waves

List of symbols

k	wave number
λ	wavelength
f	frequency
$\mathcal{T}$	period
N	number of waves
Δs	pulse width
Δt	time for one beat
Δf	frequency range $(f_1 - f_2)$
Δk	wave number range $(k_1 - k_2)$
v	wave speed
θ	diffraction angle
θ_i	angle of incidence
θ_R	angle of reflection
θ_r	angle of refraction

Problems

Here are some suggestions for problems that have to do with water waves. Observations on a natural body of water are made most effectively from a bridge or pier overhanging the water. You may observe wind-generated wave trains or pulses generated by a stone. By dipping your toe rhythmically into the water, you may be able to generate a circular wave train.

Experiments can be conducted in a bathtub or sink if natural bodies of water are not available. A pencil or comb dipped horizontally into the tub near one end can generate straight wave pulses. Dipping your finger, a pencil or a comb vertically will generate circular waves. To observe the waves, place a lamp with one shaded bulb over the bathtub so as to direct the light at the water surface and not into your eyes. You should also avoid looking at the reflected image of the bulb. Under these conditions, waves cast easily visible shadows on the bottom of the tub or on the ceiling. **Caution: You must be careful when using electricity near the bath or sink; an electrical shock from household**

current can be dangerous. Keep water away from the lamp and do not under any circumstances touch the lamp with wet hands nor while any other part of your body is wet or touching something wet.

1. Measure the speed of water waves by measuring how long they take to traverse a given distance. Describe the conditions of your observations, especially the depth of the water and the amplitude of the waves. If you observe wave trains, determine their frequency and wavelength and test Eq. 6.3b.

2. Identify the interaction(s) that are involved in the propagation of waves on a water surface.

3. Observe waves at the seashore and report qualitatively about as many of the following as you can observe.
   (a) Differences in speed of various waves.
   (b) Differences in direction of propagation.
   (c) Applicability of the superposition principle.
   (d) Effect of the depth of the water on the wave motion.
   (e) Reflection of wave fronts.
   (f) Refraction of wave fronts.
   (g) Diffraction of waves.
   (h) Transfer of energy from the waves to other systems.

4. Sand ripples are frequently observed on the ocean or lake bottom in shallow water. They are formed by the interaction of sand and water just as water waves are formed by the interaction of water and wind. Measure the wavelength of sand ripples that you observe. Comment on their propagation speed.

5. Observe reflection of water waves in your sink or bathtub. Estimate the angles of incidence and reflection as well as you can and compare your results with the law of reflection for waves.

6. Observe single-slit diffraction of water waves in your sink or bathtub. Report the slit width you found most suitable and other conditions that helped you to make the observations.

7. Several different diffraction gratings diffract water waves with a wavelength of 0.03 meter. Find the diffraction angle for a diffraction grating with a slit spacing of (a) 0.30 meter; (b) 0.10 meter; (c) 0.05 meter; (d) 0.025 meter.

8. Water waves are diffracted by a grating with a slit spacing of 0.30 meter. Find the wavelengths for the waves when the diffraction angle is (a) 10°; (b) 25°; (c) 60°.

9. Find the result of superposing the following three waves:
   wave A -- wavelength (λ) = 6 centimeters (cm), amplitude = 3 cm;
   wave B -- λ = 3 cm, amplitude = 2 cm;
   wave C -- λ = 2 cm, amplitude = 1 cm.
   Start from a point where all three waves interfere constructively; keep plotting until all three waves again interfere constructively.

10. Sound waves in air have a wave speed of 340 meters per second. Find the wavelength and wave number of the following sound waves: (a) middle C, frequency (f) = 256 per second; (b) middle A, f = 440 per second (c) high C, f = 1024 per second.

11. Use the paper strip analogue (Fig. 6.24) to study diffraction of waves. Report the wavelength, "slit" separation, and diffraction angle(s) for three different "gratings." Choose λ/d small (0.5), medium (2.0), and close to one for the three cases. (Note: one grating may give several diffraction angles, according to whether the waves from adjacent slits are 1, 2, 3, ... wavelengths out of step.) Make as many paper strips as you feel necessary to help you.

12. Use the paper strip analogue (Fig. 6.24) to study diffraction from only two slits. Measure and/or use geometrical reasoning to find the angles of diffraction amplitude maxima (constructive interference) and diffraction amplitude minima (destructive interference). Compare your results with those obtained for a diffraction grating and describe qualitatively the reasons for similarities and differences.

13. Identify one or more explanations or discussions in this chapter that you find inadequate. Describe the general reasons for your judgment (conclusions contradict your ideas, steps in the reasoning have been omitted, words or phrases are meaningless, equations are hard to follow, . . .), and make your criticism as specific as you can.

Bibliography

W. Bascom, *Waves and Beaches*, Doubleday, Garden City, New York, 1964.

C. Huygens, *Treatise on Light*, Dover Publications, New York, 1912.

U. Haber-Schaim, *PSSC Physics*, *Seventh Edition,* Kendall/Hunt, Dubuque, Iowa, 1991.

R. A. Waldron, *Waves and Oscillations*, Van Nostrand, Princeton, New Jersey, 1964.

Articles from Scientific American. Some or all of these, plus many others, can be obtained on the Internet at http://www.sciamarchive.org/.

E. D. Blackham, "The Physics of the Piano" (December 1965).

E. E. Helm, "The Vibrating String of the Pythagoreans" (December 1967).

Marin Mersenne (1588-1648). The scientific movement in France was characterized by a tradition of informal gatherings of interested scientists. One such informal group was held together by the personality of Marin Mersenne, a friar in the Palais Royale. Mersenne and his students disseminated the discoveries of Galileo, popularized the Cartesian coordinate frame, and publicized the work of such men as Pascal. Moreover, Mersenne succeeded (where Galileo had failed) in identifying the path of a falling body as having the shape of a parabola.

In this chapter we will return to the discussion of sound and light that we began in Chapter 5. This time we will describe both phenomena from the point of view of the wave model. As we already explained in Section 1.1, the history of models for light was full of controversy and the currently accepted models are still undergoing change. By contrast, the understanding of sound as wave motion dates to the seventeenth century and has advanced steadily with significant contributions by many physicists and mathematicians.

7.1 Applications of wave theory to sound

Early history. Sound and motion had been associated with one another since ancient days, but Galileo was the first person to note clearly the connection between the frequency of vibration of a sound source and the pitch of the note that was produced. This, in spite of the fact that musical instruments had existed for millennia!

Early experiments on the speed of sound. Many of the early experiments reveal a simplicity, an ingenuity, and occasionally a misconception that are charming. Thus, Galileo's idea originated in his scraping a knife at various speeds over the serrated edge of a coin. The first measurements of the speed of sound were made by timing the interval between the flash and sound of a distant gun being fired. A value of the speed very close to the present 344 meters per second was found. It was also noticed that sound travels faster in water (experiments were conducted in Lake Geneva, Switzerland) and in steel wires than in air.

Frequency of sound. Marin Mersenne, sometimes called the "father of acoustics," was the first individual to measure the frequency associated with the musical note emitted by a particular organ pipe. He tuned a short brass wire to the same musical note (and therefore the same frequency) as the organ pipe by hanging on weights to adjust the tension. Then he repeated the experiment with wire of the same material and thickness, and under the same tension, but 20 times as long. This wire vibrated so slowly that he could count ten vibrations per second. Because the material, thickness and tension of the two wires were identical, the wave speed along the two wires was the same. The relationship between frequency and wavelength (Eq. 6.3, $f = v/\lambda$) shows that, if v is constant, the frequency is *inversely proportional* to wavelength. (To review inverse proportions see Appendix A.2.) Thus Mersenne concluded that the original short wire (and the organ pipe) had executed 200 vibrations per second, 20 times as many as the long wire.

Medium for sound propagation. One of the big questions that had to be resolved was whether sound needed a medium. Vacuum pumps had been invented about the middle of the seventeenth century, and it was a simple matter to suspend an alarm clock or a bell in a jar to be evacuated. Unfortunately, a decade was required before Robert Boyle (1627-1691) successfully observed that the alarm in the jar could not be heard outside when the jar did not contain air. Other investigators had arrived at the contrary conclusion, perhaps because they had failed to remove the air completely enough, or perhaps because the bell's support conducted the sound to the observers.

By the time of Isaac Newton, the wave model for sound as vibration of an elastic medium was generally accepted. In fact, it was so well accepted that the differences between sound and light, which we described in Section 5.1, convinced Newton that light could not also be a wave phenomenon (see Section 7.2)!

Speed of sound. In Section 6.1 we described the conditions for wave motion in terms of the oscillator model for the wave medium. The two key factors were the inertia of the oscillators making up the medium and the interaction among them. The gasbag model for air and the MIP model for solids and liquids represent these materials as composed of subsystems that have inertia and interact with one another. Sound waves, therefore, propagate in all materials, but with a predicted speed that is high if the oscillators have low inertia and/or strong interaction and low if the conditions are opposite. Values for the speed of sound in various materials are listed in Table 7.1.

TABLE 7.1 SPEED OF SOUND IN SOLIDS, LIQUIDS, AND GASES

Material	*Speed (m/sec)*
metals:	
aluminum	*5100*
brass	*3500*
copper	*3560*
gold (soft)	*1740*
iron	*5000*
lead	*1230*
brick	*3650*
glass	*5000*
marble	*3800*
paraffin	*1300*
rubber	*54*
liquids:	
alcohol	*1240*
water	*1460*
gases (room temperature):	
air	*344*
carbon dioxide	*277*
helium	*960*
hydrogen	*1360*

According to the MIP model, hard materials, in which there is strong interaction between oscillators, should exhibit a higher sound velocity than "soft" materials. You can see a trend compatible with your expectation; rubber, lead, paraffin, and water have a relatively low sound speed, while glass, iron, and aluminum have a high sound speed.

Speed of sound in gases. Gases are more difficult to include in the comparison, because both the interaction and the inertia in these low-density materials are much smaller than in liquids and solids, and these two differences may compensate for one another. From the fact that the sound speed in gases is lower than that in solids or liquids, you can conclude that the reduced interaction strength is more significant than the reduced inertia. This analysis of sound speed in gases is an example where the model does not lead to an unambiguous prediction, but where the model and experimental data may be combined to yield more insight into the properties of matter.

Frequency of musical notes. The wave model explains musical notes of different pitch as vibrations of different frequency. This relation was first investigated quantitatively by Mersenne. The presently accepted standard frequency is 440 vibrations per second for the "middle A" note. The entire musical scale is divided into octaves, which are two notes with a frequency ratio of two to one. Thus, various A notes have vibration frequencies of 110 per second, 220 per second, 440 per second, 880 per second, and so on. In Western music since about 1800 the octave interval is generally divided into twelve notes ("semitones"). The frequency ratio of adjacent semitones is slightly less than 1.06. In other words, each semitone has a frequency almost 6% larger than the next lower semitone. The notes in one octave and their frequencies are listed in Table 7.2. All frequencies in the table are multiples of the standard A-440 frequency. You can calculate the frequencies of the corresponding notes in higher or lower octaves by successively doubling or halving the frequencies in the table.

TABLE 7.2 PROPERTIES OF SOUND WAVES FOR MUSICAL NOTES

Note	*Frequency (f, /sec) (approximate)*	*Frequency ratio to C note (approximate)*	*Wavelength in air (λ, m) (= 344/f, approx.)*
C (middle C)	262	1/1	1.31
C# = D^b	277	-	-
D	294	-	-
$D^\#$ = E^b	311	-	-
E	330	5/4	1.04
F	349	4/3	0.99
$F^\#$ = G^b	370	-	-
G	392	3/2	0.88
$G^\#$ = A^b	415	8/5	0.83
A (standard)	440	5/3	0.78
$A^\#$ = B^b	466	-	-
B	494	-	-
C	523	2/1	0.66

Equation 7.1

$$v = f\,\lambda,$$

$$or,\ \lambda = \frac{v}{f}$$

$$or,\ f = \frac{v}{\lambda}$$

EXAMPLE:
$v = 344$ *m/sec,*

$f = 262$ */sec*

$$\lambda = \frac{344}{262}\,m = 1.31\ m$$

Wavelengths of sound waves. Air is, of course, the most important medium for the transmission of sound on earth. You can calculate the wavelengths, λ, in air of musical notes from their frequency (f, Table 7.2, second column) and the known sound speed, v, in air (344 m/sec from Table 7.1), by using the equation v = f λ in the form λ = v/f (Eq. 7.1, from Eq. 6.3b). The results are included in Table 7.2, fourth column. It is clear that audible sound waves, especially the ones used in speech, have a wavelength comparable to the size of the human body and to objects in our environment. This result (wavelengths of a few feet) is not surprising—as explained above (Section 6.2), the lengths of organ pipes range from a few inches to many feet, which is also the approximate size of the wavelengths of the notes they produce.

The magnitude of wavelengths of audible sound in air explains, in the context of the wave theory, why it is impossible to form a sharp acoustic image of the placement and shape of primary sound sources or reflectors. We pointed out in Section 6.4 that obstacles whose size is comparable to the wavelength diffract waves most strongly. Diffraction by persons, furniture, doors, and buildings, therefore, bends the sound waves so much that their direction and intensity is related only remotely to the placement of the primary sound sources and the reflecting surfaces. Sound transmits certain information about the primary source, for instance, intensity, pitch, and duration, but no sharp image of the location of the sound sources. Only in a clear space, and with the help of both ears, which receive somewhat different information (stereophonic), can we determine the position of sound sources in even an approximate way.

Musical instruments. In Section 6.2 we explained standing waves in an organ pipe and on a violin string as standing waves on tuned systems. The musical octave is simply related to the musical intervals between notes that can exist in a tuned system. For illustration, the various notes generated by standing waves in an organ pipe of 0.657-meter

TABLE 7.3 STANDING WAVES IN AN ORGAN PIPE WITH LENGTH (L) = 0.657 METER

Wavelength (meters) (λ)	Frequency (/sec) ($f = v/\lambda = 344/\lambda$)	Note on musical scale
1.31=2 L	262	C
0.66=2/2 L	524	C'*
0.44=2/3 L	785	G'*
0.33=2/4 L	1047	C''*
0.26=2/5 L	1309	~E''†
0.22=2/6 L	1571	G''*

* C' indicates a note exactly one octave above middle C with frequency = 2 x 262 = 524; C'' indicates a note exactly two octaves above middle C with frequency = 4 x 262 = 1048. G', G'' and E'' are defined similarly with respect to G and E in Table 7.2.

† The note with frequency of 1309 corresponds only approximately with E''.

length are listed in Table 7.3. The lowest frequency wave (longest wavelength) has nodes at both ends of the pipe with half the wavelength equal to the length of the pipe, or a full wavelength equal to 2L (first line of Table 7.3). The frequency is determined from Eq. 7.1, $f = v/\lambda = 344/1.32 = 262$ /sec. This wave is known as the fundamental; it has the lowest frequency possible on this system. Other waves (overtones) must also have nodes at the ends of the pipe, but additional half wavelengths can be fitted within the length of the pipe; this determines the remaining wavelengths (and higher frequencies) in Table 7.3, as explained in Sect. 6.2 (Eqs. 6.4 and 6.5).

Stringed instruments. We will explain stringed instruments (such as the guitar and violin) using the wave model. A single vibrating string does not transfer energy and sound effectively to the air; therefore, all stringed instruments require amplification and/or a well-designed sound chamber (the instrument's hollow body), which acts as a coupling element (Section 4.3) to the air. The sound chamber is passive in the sense that it does not affect the transfer of energy, but it is very important in determining the "quality" of the sound we hear. The best instruments are made from wood; the specific characteristics of the wood and its finish (the surface of which actually transfers the sound to the air) are critical.

In a guitar or violin, the lengths of all the strings (and thus the wavelengths of the sounds traveling along the strings) are determined by the length of the instrument and, therefore, are all the same. For the instrument to be able to produce a sufficiently wide range of pitches, the various strings, played at full length, must, therefore, vibrate at different frequencies. Waves of equal wavelengths but different frequencies must travel at different wave speeds along the various strings (Equation 7.1, $v = f\lambda$). Differing speeds means the various strings must have different inertia (weight or density) and/or a different strength of interaction (tension) along the string.

The first column of Table 7.4 lists the notes sounded by the six strings of a guitar when vibrating at full length; the second column lists the frequencies of these notes. To "tune" the strings so they vibrate at

TABLE 7.4 GUITAR WITH STRINGS 0.65 METER LONG (= 1/2 λ)

Note sounded by string (at full length)	Frequency (f, /sec) (from Table 7.2)	Wave speed on string (v, m/sec) (= f λ) (= f x 1.3m)	Wavelength in air (λ, m) (= v_{air}/f)
E	82.5	107	4.17
A	110	-	-
D	147	-	2.34
G	196	-	-
B	247	-	-
E	330	429	1.04

exactly the correct frequency, the player turns the pegs, slightly adjusting the tension and thus the wave speed.

We can use the information in Table 7.4 to find the actual speed of the waves on the strings. The strings are all 0.65m long, so there must be nodes at both ends; thus half the wavelength must be 0.65m, and λ = 1.30 m. Putting this and the frequency into Equation 7.1 yields the wave speed on the string (Table 7.4, third column). Alternatively, using the frequency plus the speed of sound *in air* in Eq. 7.1 gives the wavelength in air (fourth column). Note that it is easy to forget that the wave speed (v) in Eq. 7.1 depends on the medium (sound or air). You should be careful to use the appropriate value, depending upon the medium (the substance that is actually vibrating and carrying the wave).

Other sound phenomena. We began the discussion of sound in Chapter 5 with the unexpressed operational definition of sound, "sound is what people can hear." It is now appropriate to redefine sound with a formal definition, as displacement waves in a medium, and thereby to extend the concept of sound beyond the limitations of the human ear. As a matter of fact, the human ear is capable of detecting sound waves only between frequencies of approximately 20 and 20,000 vibrations per second, with a great deal of variation among individuals. Lower-frequency vibrations are sensed as rapid knocking, while higher-frequency vibrations are not detected at all, except possibly as pain if they are very intense. Nevertheless, these other waves do occur naturally and/or have been exploited technologically.

Ultrasonics. Ultrasonics refers to sounds with frequencies too high for the human ear. Therefore, ultrasonic sound waves have a much shorter wavelength than audible sound, only about 0.01 meter or less. Ordinary size obstacles therefore diffract ultrasonic sound waves much less than audible sound; hence ultrasonic sound waves can be directed into narrow beams that are reflected by environmental objects. The reflected beam furnishes information about the position of the reflecting object.

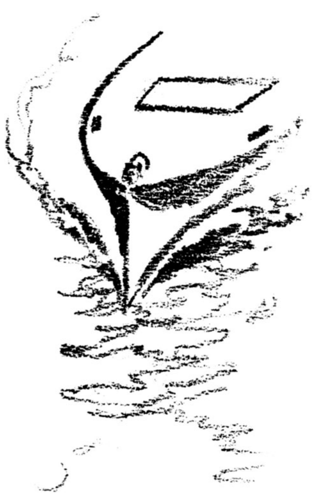
Sonar is a method for locating objects under water using ultrasonic sound waves. A high frequency sound source emits wave pulses at a frequency of 20,000 vibrations per second or more; a detector records the reflected pulses (echoes). The relative position (direction and distance) of the reflecting object is determined from the direction of the reflected pulse and the time delay of its arrival after the original pulse was emitted. Sonar depth gauges, which measure the distance to the ocean floor by the time delay of reflected pulses, are now standard equipment on many pleasure boats and commercial craft.

In an industrial application of the sonar principle, sound with several million vibrations per second is used to locate flaws in steel pieces, rubber tires, and so on. The flaw is an irregularity that reflects sound waves and can thereby be detected.

The sonar principle has also been applied in many beneficial ways to health care, most notably to form images of a developing human fetus within the mother's womb. The baby's tissues and bones reflect the sound waves, and computer-assisted detectors can then, amazingly enough, produce an image from the reflected waves and display it on a conventional TV monitor. Naturally the potential effects of the sound waves on the human body must be investigated carefully and, so far, such effects have been found to be negligible.

The bat is an unusual mammal that can use the sonar principle to locate objects and avoid obstacles in dark spaces (such as caves and belfries). A bat can make sounds with frequencies close to 100,000 vibrations per second. Bats use these sounds to perform amazing feats locating tiny insects (their food) while flying in pitch darkness.

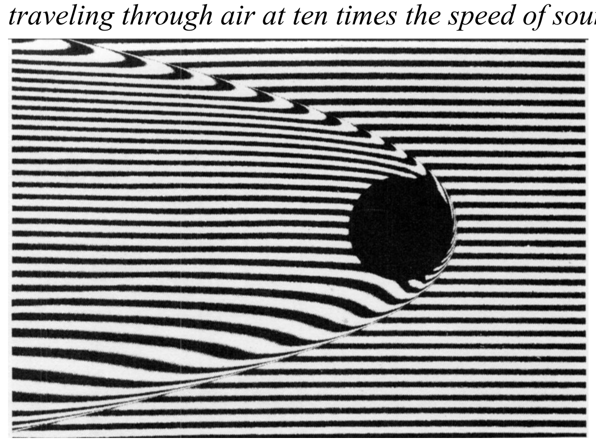
Sub-audible waves. At the other end of the sound spectrum from ultrasonic waves are waves with sub-audible frequency. Most interesting to the scientist are seismic waves, which are generated by the movement of large bodies of rock during earthquakes. The frequencies of seismic waves are in the range of a few vibrations per minute (0.1 per second). There are two kinds of seismic waves, which differ in the direction of the oscillator displacement relative to the direction of propagation. One kind, called the primary wave, travels about 6500 meters per second in the earth's crust, twice as fast as the other kind, called the secondary wave.



Seismic waves are the best source of information about the interior of the earth. They are refracted inside the earth because their speed in various layers is greater or less than it is at the surface. The earth therefore acts like a huge, complicated lens whose properties are inferred from the geographic distribution of seismic waves emitted in earthquakes. One inference is that the material changes abruptly at a depth of about 50 kilometers. This change, which defines the boundary of the earth's crust, is called the Mohorovicic discontinuity ("Moho" for short).

Shock waves. The final item we will take up in this section is shock waves, which are a form of sound with extremely large amplitude and very sudden onset. Whenever an object moves with supersonic speed (faster than the speed of sound in the surrounding air or other medium), the air is displaced very abruptly. What happens then is analogous to what happens at the bow of a speedboat that pushes the water aside

Figure 7.1 A shock wave created by a plastic sphere traveling through air at ten times the speed of sound.



suddenly. The sudden displacement of the air by the moving object cannot communicate itself to other parts of the air in the form of sound waves, because sound travels too slowly. For example, at the front of the moving object, the sound can't get "ahead" of the object because the object itself is moving faster than sound.

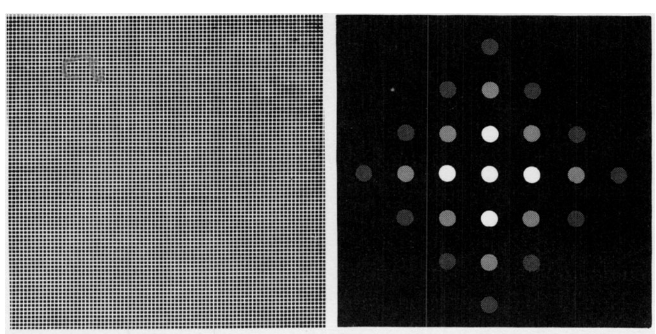
Consequently, there is a very large change of air pressure, the ordinary wave model breaks down, and the frequently destructive shock wave is formed (Fig. 7.1). Supersonic airplanes generate shock waves (sonic boom) in air very much in the way the speedboat generates shock waves on the water surface. The boom is caused by the sudden increase in air pressure. Explosions also generate shock waves. The very hot material near the site of the explosion expands into the surrounding material with a speed faster than the speed of sound in that material.

7.2 Application of the wave model to light

The wavelength of light. Your observation of an interference pattern when you looked through a piece of cloth at a distant light source (Section 5.1) seems to be understandable only with a wave model for the light (Figure 7.2). When you apply the wave model to your everyday experience with light, you conclude, from the absence of noticeable diffraction under ordinary circumstances, that the wavelength of light must be much smaller than the size of the objects around you. Only when you looked through finely woven fabric were the effects of diffraction noticeable, and even then they were quite small.

You can understand the appearance of the interference pattern by thinking of the threads in the fabric as forming two diffraction gratings, one with its slits and barriers at right angles to those of the other one.

Figure 7.2 A handkerchief serves as diffraction grating (Fig. 5.3).
(a) Thread pattern of a handkerchief.
(b) Diagram of the interference pattern from a distant lamp observed through a handkerchief.



(a) (b)

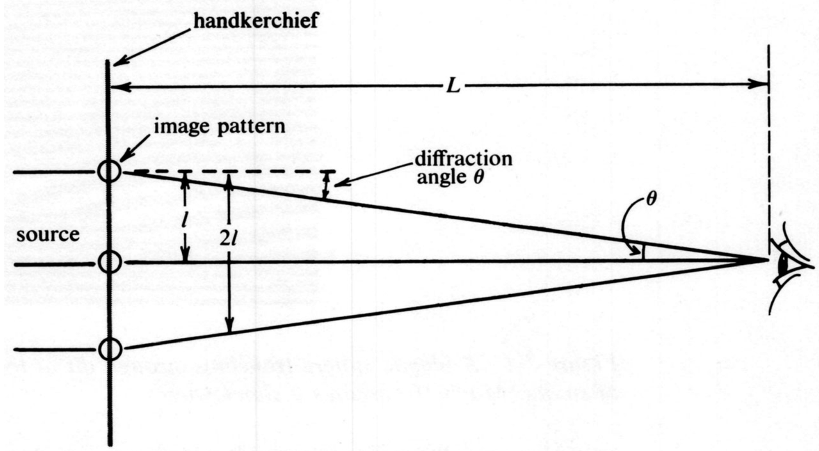


Figure 7·3 Measuring the wavelength of light with a handkerchief held at arm's length, a distant light source, and a ruler (example included for illustration).

(1) Measure arm length: $L \approx 70\ cm = 700\ mm$.

(2) Measure the spacing of the three bright image spots at a distance of one arm's length: $2\ l \approx 3\ mm$.

(3) Use Eq. A·6 to calculate:

$$\text{sine}\ \theta \approx \frac{l}{L} \approx \frac{1.5}{700} \approx 2 \times 10^{-3}$$

(4) Measure the "slit distance" of the handkerchief, which has three threads per millimeter (use a magnifier):

$$d \approx 1/3\ mm \approx 0.3\ mm = 3 \times 10^{-4}\ m$$

(5) Calculate the wavelength (Eq. 6·13):

$$\lambda = d\ \text{sine}\ \theta \approx 3 \times 10^{-4} \times 2 \times 10^{-3}\ m = 6 \times 10^{-7}\ m$$

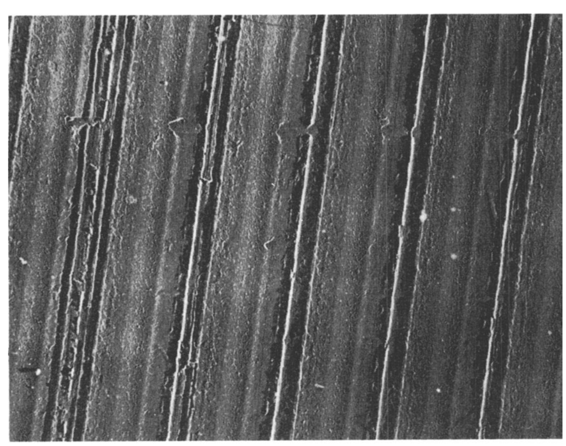


Figure 7.4 Electron microscope photograph of a diffraction grating.

The first machines for making the slits in a grating were designed and built by Henry Rowland of Johns Hopkins University at the end of the nineteenth century; his gratings, made by scribing many precise scratches on metal or glass, were expensive and prized scientific tools. Nowadays, very inexpensive gratings are manufactured by impressing the rulings on a sheet of plastic, in much the same way that CDs or auto parts are stamped out from a master mould.

The interference of the light diffracted by the vertical threads produces images of the source displaced in the horizontal direction. The horizontal threads diffract the light to produce images that are displaced in the vertical direction. The combination of both then results in the checkerboard array of images that is observed (Fig. 7.2). How a simple measurement can be used to calculate the wavelength of visible light is explained in Fig. 7.3. The wavelength is indeed very short, only about 6 x 10^{-7} meters.

Diffraction gratings. Diffraction gratings for the study of light have to be made with a spacing between slits that is comparable to the wavelength. Then the light is diffracted at angles that can be observed easily. A commercial diffraction grating is a transparent sheet with many narrow scratches on its surface (Fig. 7.4). The scratches, which are too small to be seen, are slight obstacles to the propagation of light. The narrow regions between the scratches therefore act as narrow slits. Most of the light incident on the grating passes through unaffected. A small portion of the light, however, is diffracted by the many slits and emerges traveling in a direction at an angle to the incident light. As was shown in Section 6.4, Eq. 6.13 relates the diffraction angle to the distance between slits and the wavelength of the light. Light of

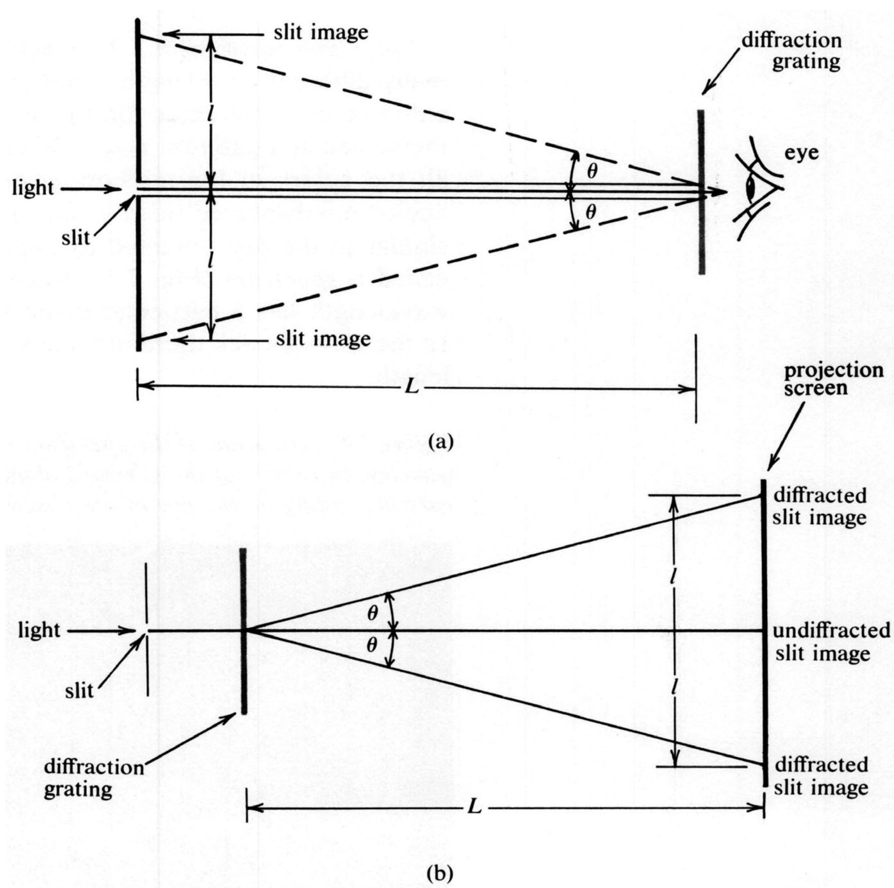


Figure 7.5 Observation of the light that comes from a slit and passes through a diffraction grating.

(a) Viewing the diffracted light directly.

(b) Projecting the diffracted light on a screen.

Equation 6·13

wavelength (meters) λ
distance between slits (meters) d
diffraction angle θ

$$\text{sine } \theta = \frac{\lambda}{d}$$

Equation 7·2

(Symbols defined in Fig. 7·5.)

$$\text{sine } \theta = \frac{l}{\sqrt{l^2 + L^2}} \approx \frac{l}{L}$$

Equation 7·3

$$\frac{\lambda}{d} = \frac{1}{\sqrt{1^2 + L^2}} \approx \frac{l}{L}$$

short wavelength, therefore, is diffracted through a smaller angle than light of long wavelength.

To observe the diffracted light, you may either look through the grating at the light source or let the diffracted light impinge on a screen at a substantial distance behind the grating (Fig. 7.5). A shield with a slit is usually placed in front of the grating to act as a narrow rectangular light source whose diffracted images can be recognized easily by their shape. You can use the measurements to calculate the diffraction angle (Eq. 7.2) or you can calculate the wavelength directly (Eq. 7.3).

If the wavelength is very much smaller than the grating spacing, then the diffraction angle is very small and you cannot observe the diffracted light separately from the undiffracted light because the two images of the slit overlap. For good observations, the grating spacing must be almost as small as the wavelength of light. The manufacture of such gratings clearly requires precision apparatus.

Color and wavelength. If a beam of white light, which includes light of many different wavelengths strikes a diffraction grating, the various wavelengths emerge at different angles and thus strike a screen (as shown in Fig. 7.5) at different locations. What your eye observes on the screen, however, is a display of all the colors of the rainbow side by side (as shown in Fig. 7.6). Such a display of light, similar to the one obtained by Newton with a prism (Section 5.2), is called a *spectrum*. Each portion of the light with a single wavelength has a particular color and is called monochromatic light. In the wave model, therefore, *color of light is associated with its wavelength.*

The wavelength of visible light ranges around the value we reported above from the crude experiment with the handkerchief. The association of color and wavelength is given in Table 7.5. As shown in the Table, the wavelength range of visible light is actually quite narrow (from 4 to 7 x 10^{-7} m). This seems somewhat paradoxical: the colors detectable by the human eye seem, intuitively, to span a huge range (think about the number of colors available in a paint store, or, even more impressive, the complex shades and hues discovered by the Impressionist painters), yet this extraordinary complexity is confined within such a narrow range of wavelengths! However, on reflection,

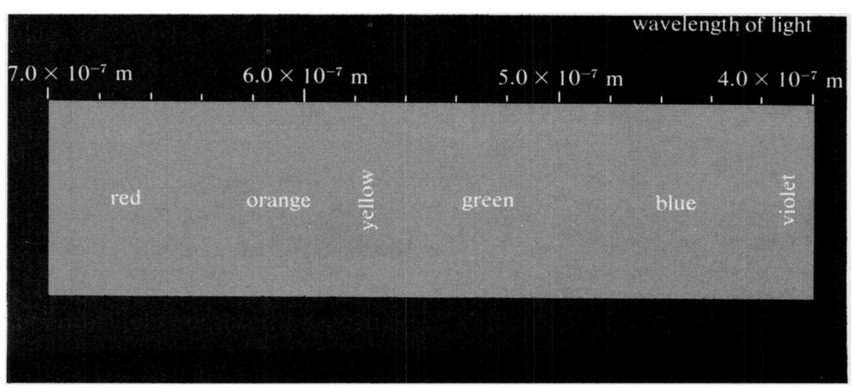


Figure 7.6 Diagram of the spectrum of light, with an indication of the wavelength ranges of the various colors. The transitions are gradual and vary depending on the eye of the observer

TABLE 7.5 WAVELENGTH AND COLOR OF LIGHT (APPROXIMATE)

Color	*Wavelength range (m)*
violet	*4.0 to 4.2 x 10^{-7}*
blue	*4.2 to 4.9 x 10^{-7}*
green	*4.9 to 5.7 x 10^{-7}*
yellow	*5.7 to 5.8 x 10^{-7}*
orange	*5.8 to 6.4 x 10^{-7}*
red	6.4 to 7.0 x 10^{-7}

Ole Roemer (1644-1710) was a prominent Danish scientist who served as a member of the French Academy and as the tutor of the son of King Louis XIV. With the revocation of the Edict of Nantes in 1681, Roemer, like Huygens and others prominent in the French Academy, fled France for the safety of Protestant Northern Europe.

Equation 7.4 (Roemer's measurement of the speed of light in 1676)

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

$$\Delta s = diam., \, earth\, orbit = 3 \times 10^{11} m$$

$$\Delta t = 22 \min. = 1.3 \times 10^{3} \sec$$

$$v_{light} = \frac{3 \times 10^{11} m}{1.3 \times 10^{3} \sec} = 2.3 \times 10^{8} m / \sec$$

Modern value:
v_{light}= 3.0 x 10^8 m/sec

there is a refreshing lesson here; our eyes (and brains) indeed are capable of quite extraordinary feats as detectors of light, able to easily distinguish colors that have almost identical wavelengths (or slightly different mixtures of wavelengths). Modern spectroscopes and other instruments are extraordinarily good at spreading light out and detecting even the faintest components, but our eyes are also extremely capable. The human visual system (the eye and brain) is also capable of making extraordinarily fast judgments in real time; while the best optical instruments probably could match or exceed the human eye in the narrowly-defined task of distinguishing between slightly different wavelengths; there would be no contest with regard to speed.

We can also think of the differences among the various colors of light in terms of their frequencies. The frequency of visible light (calculated from f = v/λ) ranges from about 1.3 to 2.3 x 10^{15}! This is an extraordinarily high frequency. The complex information about colors that can be conveyed by light is a example of the huge information-carrying ability of a wave with such a high frequency. It is indeed generally true that higher frequency waves can carry more information. Another example of this is in the capacity of fiber optic cables (which use light) to convey information, which far exceeds the capacity of coaxial cable or ordinary telephone wires (both of which use much lower-frequency waves in the radio range).

By using energy detectors other than the human eye it is possible to identify diffracted radiation of shorter and of longer wavelength than visible light. This radiation is called *ultraviolet* light and *infrared* light, respectively. In other words, the implicit operational definition for light, "radiation detected by the human eye," should be extended to forms of radiation that are not detected by the eye, but are functionally very similar to visible light.

The speed of light. Ancient philosophers speculated about the speed of light and variously held the opinion that light propagated with a finite speed and that it propagated instantaneously. Galileo made the first attempt to measure the speed of light by having two distant observers flash lanterns back and forth. The experiment failed because the time required by the light was much less than the reaction time of the participants.

Roemer's measurement. Ole Roemer made the first successful measurement of the speed of light in 1676. He studied the revolution of Jupiter's satellites (the four "Galilean moons," discovered by Galileo in 1610). Roemer noticed something strange: the moons' orbital periods all became gradually shorter while the earth was approaching Jupiter and longer while the earth was moving away from Jupiter. Roemer figured out that these changes must be due to the fact that light did not travel at infinite speed. In fact, light took some time to travel from Jupiter to the earth, and this delay would gradually become shorter (or longer) when the earth was approaching (or moving away from) Jupiter. Roemer measured the maximum time difference in the revolution periods to be 22 minutes (1.3 x 10^3 sec), during which time the light would have to cross the earth's orbit, a distance that was then thought to be 3.0 x 10^{11} meters. Even though Roemer's result (Eq. 7.4) is considerably lower than the presently accepted value, it is a truly remarkable achievement because it occurred only a century after the planetary model for the solar system was introduced by Copernicus, only 66 years after Jupiter's moons were discovered by Galileo, and at a time when the diameter of the earth's orbit around the sun was not known accurately.

Foucault's measurement. Modern methods for measuring the speed of light basically make use of Galileo's concept but substitute a mirror rotating at a known high speed (Foucault's contribution) for the man flashing the lantern. The rotating mirror flashes a beam of light to a distant mirror, which returns the light with a delay equal to the time required for the light to travel to and from the distant mirror. Depending on the time delay, the rotating mirror has assumed a new position, which reflects the returning light to a detector. Because the mirror rotates at high speed, even the short travel time of the light flash finds the mirror in a measurably changed position. This change in the mirror's position is compared with the known speed of rotation to yield the travel time. The speed of light is then found by dividing the distance the light traveled by the time. (Table 7.6).

Jean Bernard Lion Foucault (1819-1868) studied medicine before changing to physics. In 1851 his celebrated Foucault pendulum experiment demonstrated the earth's rotation relative to the fixed stars. In later years, he invented the gyroscope and made a determination of the velocity of light by using a revolving mirror.

It is ironic that modern methods for measuring the dimensions of the solar system represent a reversal of Roemer's procedure. Now that the speed of light is known accurately from terrestrial measurements, the travel time of light to other planets is observed (as it was by Roemer) and the distance to them is calculated.

Speed of light in water. Foucault also measured light's speed in water and found that it travels considerably *slower* than in air. This behavior is contrary to the behavior of sound, which travels *faster* in dense materials than in air (Table 7.1). For this reason, as we explain below, Foucault's measurements of the speed of light were very significant in the history of the theory of light; Foucault's measurements provided a critical test for models of light and dramatized the inadequacy of the wave model. Later measurements confirmed Foucault's results for water and found that light also travels more slowly in glass than in air (Table 7.6).

TABLE 7.6 SPEED OF LIGHT

Material	Speed (m/sec)
vacuum	3.0 x 10^8
air	3.0 x 10^8
glass	1.9 x 10^8
water	2.3 x 10^8

The ray model and the wave model for light. The ray model described in Section 5.2 was based on a set of assumptions that were not further justified. It was sufficient that they were successful in explaining the observed properties of light, such as formation of shadows, operation of lenses, combination of colors, and so on. The model described but did not explain refraction (Fig. 5.16, Assumption 5) or the difference among monochromatic rays of various colors. Furthermore, it did not specify how a single ray might be separated from a light beam; that is, the ray was a formal concept in the model and did not have an operational definition.

Isolation of a light ray. Since the ray model says that light is composed of rays, you may well be curious to see a single such ray. Suppose we attempt to isolate a single ray as follows: We place an opaque shield in front of a light source and puncture it. Through the tiny hole, a

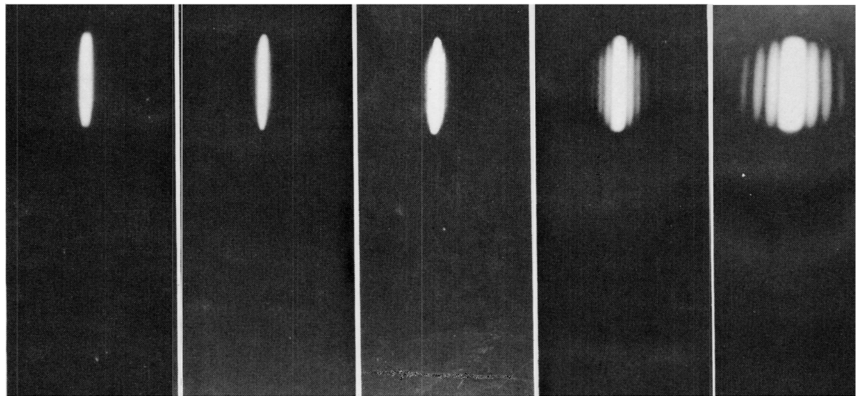


Figure 7.7 The attempts to isolate a single light ray by passing light through a narrow slit fail. The slit widths (in order from left to right) are 1.5, 0.7, 0.4, 0.2, and 0.1 millimeters, respectively.

slim beam of light passes. Now we make successively smaller and smaller punctures in the shield. The ray model predicts that we should get thinner and thinner shafts of light. However, the physical world doesn't always act the way we expect: Figure 7.7 shows what actually happens.

The light actually *spreads out* as the hole is reduced below a fraction of 1 millimeter in width (Fig. 7.7)! This behavior, especially the pattern of light and dark fringes in the photo with the narrowest slit (on the far right) is very mysterious; it doesn't fit at all naturally with the ray model. You may think that the spreading out of the light could be explained within the framework of the ray model by reflection or scattering of the light in some way from the edges of the slit; scientists, including Newton, indeed used the ray model to construct such explanations. However, any explanations based on the ray model simply cannot explain the pattern of dark and light fringes that appear as the slit gets narrow. In fact, the narrower the slit, the wider and more pronounced the fringes become, and there isn't any way to combine rays of light in such a way as to cancel themselves and thus produce a dark fringe.

On the other hand, this phenomena is very reminiscent of what we observed with waves in Chapter 6: waves naturally spread out or diffract when they pass through narrow openings (Section 6.4); furthermore, waves can easily cancel one another, as in destructive interference (Section 6.2, Fig. 6.11). In addition, the two-hole interference pattern (Figure 6.21) had certain locations where the waves always cancelled one another out; this would be a natural way to explain the dark fringes. Finally, Huygens' Principle (Section 6.3) applied to waves striking a diffraction grating (Section 6.4) predicted that, as the angle of diffraction changed, the waves cancelled and reinforced and cancelled and reinforced (Figure 6.24); this would seem likely to produce a pattern of dark and light fringes.

Limitation of the ray model. Evidently the ray model is limited. When experiments are pushed beyond the limits of this model, it breaks down. The wave model is suggested by the diffraction of the single slit (Fig. 7.7), and by the interference pattern seen through the handkerchief (Fig. 7.2). As we will show below, it is a better model for light than the ray model. In other words, light beams are better represented as packets of

Equation 6·14

angle of incidence θ_i
angle of reflection θ_R

$$\theta_i = \theta_R$$

Equation 6·17b

angle of refraction θ_r
speeds of light v_i, v_r

$$\frac{\text{sine } \theta_i}{\text{sine } \theta_r} = \frac{v_i}{v_r}$$

Equation 5·2

indices of refraction n_i, n_r

$$n_r \text{ sine } \theta_r = n_i \text{ sine } \theta_i$$

light waves than as bundles of rays. The diffraction visible in Fig. 7.7 may be considered a consequence of the uncertainty principle for waves (Section 6.2). Attempts to localize the wave packet (that is, make it thinner from side to side) require a mixture of a broader range of wave numbers and wavelengths (that is, some of the light has a larger wavelength and thus spreads out sideways after leaving the slit). Within the limits of the uncertainty principle, or in the absence of diffraction, the ray model is satisfactory.

Adequacy of the wave model. Does the propagation of wave packets correctly explain Assumptions 1 to 5 (Fig. 5.16) about light rays? The answer is that it does, in view of Huygens' principle and the laws of reflection and refraction of waves (Eqs. 6.14 and 6.17b). The observed reflection and refraction of light corresponds directly with the observed behavior of water waves. In fact, the index of refraction of a material (Eq. 5.2) acquires a dramatic new significance in the wave model: it is the ratio of the speed of light waves in air to the speed of light waves in that medium (Example 7.1). With this new insight, a table of light speeds in various media can be constructed from that of indices of refraction (Table 5.1), with no measurement other than the speed of light in air (Table 7.5).

EXAMPLE 7.1

From Equation 5.2:

$$\frac{\text{sine}\ \theta_i}{\text{sine}\ \theta_r} = \frac{n_r}{n_i} \quad (1)$$

and from Equation 6.17(b)

$$\frac{\text{sine}\ \theta_i}{\text{sine}\ \theta_r} = \frac{v_i}{v_r} \quad (2)$$

Let the incident medium be air. Then $n_i = 1$ (from Table 5.1), and $v_i = v_{air}$, $n_r = n_{medium}$, and $v_r = v_{medium}$.

Putting together Eqs. (1) and (2):

$$n_{medium} = \frac{v_{air}}{v_{medium}} \quad (2)$$

However, there is a contradiction lurking in the background: the speed of sound is *greater* in water and glass than in air, but the speed of light is *less* in water and glass than in air. This may seem like a small detail, and in the 1700s it was. But much later, in the late 1800s, after the wave model for light had become very well accepted, scientists recognized that this contrast between sound and light pointed to a very serious limitation of the wave model for light. In fact, scientists' attempts to use their experience with sound (and other) waves to identify the *medium* for light waves generated many other contradictions and problems that were only resolved with Einstein's revolutionary theory of relativity. We will explain this more fully below in Section 7.3.

"Are not all Hypotheses erroneous, in which Light is supposed to consist in Pression or Motion, propagated through a fluid Medium?"

Isaac Newton
Opticks, 1704

Early history of models for light. The role of Isaac Newton in the development of models for light makes a remarkable chapter in the history of science. It is clear from Newton's writings that he understood Huygens' wave theory and that he was informed, through his own experiments and those of others, of the properties of light known in his day. These properties included the speed of light as measured by Roemer, the diffraction of light by a thin slit, the interference of light to form colors by multiple reflection from thin films (for example, soap bubbles), the association of color with wavelength (ultraviolet the shortest and red the longest), as well as the phenomena on which Newton based his formulation of the ray model.

Newton's rejection of the wave model. Newton summarized all these data in a set of rhetorical questions that defined the wave model for light. Included in his reasoning was the existence of a medium (aether) whose properties he estimated by assuming that the light waves were pressure waves in aether analogous to sound waves in air. Newton gave three principal reasons for rejecting the wave model.

First, Newton expected that light waves would be diffracted more extensively than observations showed. He dismissed the diffraction that had been observed as being too small to arise from the interference of waves and ascribed it instead to a repulsive interaction with the edges of the slit.

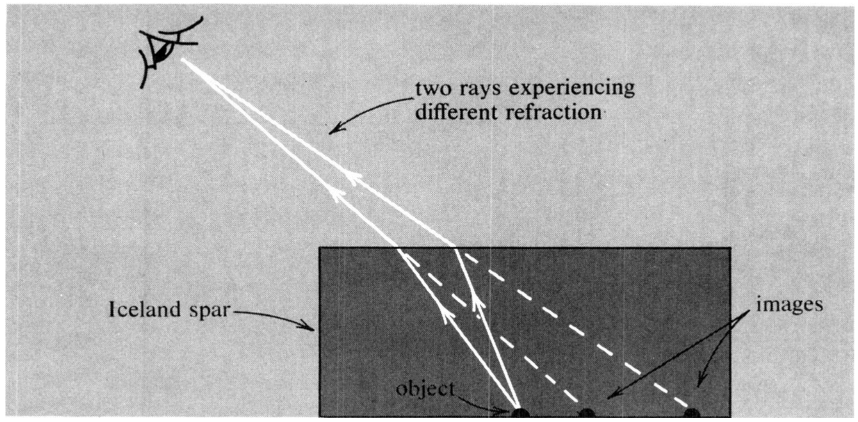


Figure 7.8 Double refraction of light by the mineral Iceland spar. Two images are seen. The distance between images depends on the viewing angle and the thickness of the mineral specimen.

Second, Newton found that the minerals Iceland spar and crystal quartz could split a light beam into two beams refracted through different angles (double refraction, Fig. 7.8). This was incompatible with Newton's concept of a wave unless the crystal itself were to modify the light wave. But one of Newton's fundamental assumptions was that the properties of light other than speed and direction were determined by the source and not by the media traversed.

Third, Newton rejected the aether concept because the very large speed of light required properties (extremely low inertia, strong interaction) that seemed unphysical. Furthermore, Newton was unable to find any reason for its existence other than that it could serve as a medium for the propagation of light.

"Are not the rays of Light very small Bodies emitted from shining Substances? For such Bodies will pass through uniform Mediums in right Lines...."

Isaac Newton
Opticks, 1704

Newton's corpuscular model. Having thus demolished the wave model, Newton proposed his own. Upon his belief that light consisted of "very small Bodies," or corpuscles, Newton built the corpuscular model for light. Besides particle structure, other features of this model were an association of particle size with color, an interaction that accelerated and deflected corpuscles falling on a dense medium (refraction), and an MIP model for matter in which the light corpuscles were easily emitted and absorbed by matter particles. The corpuscular model for light rays answered Newton's three objections to the wave model. The corpuscles can interact-at-a-distance and be deflected, but are not diffracted. The double refraction by Iceland spar and quartz crystals was explained by ascribing a shape to the corpuscles; depending on how the corpuscles were aligned relative to the micro-domain structure of the crystals, they would be deflected through different angles (Fig. 7.9). And the aether was unnecessary.

These compelling arguments show how, in the formulation of scientific models, it is usually necessary to make a compromise: some parts are more satisfactory and other parts are less satisfactory. The greatest

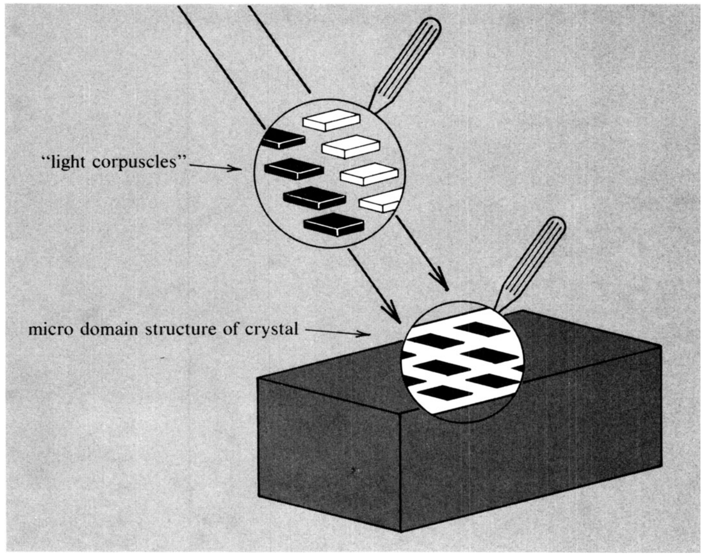


Figure 7.9 Working model for Iceland spar to illustrate Newton's interpretation of double refraction. Newton ascribed double refraction to the shape of the "light corpuscles," which could be oriented parallel or at right angles to structures in the crystal. (The details of this model were invented by the author.)

triumph of wave theory was the law of refraction, purchased at the cost of an aether. Partly because of his idea that refraction of rays was deflection of corpuscles, Newton was not willing to pay the price.

Huygens' preference for the wave model. Huygens, however, was willing to pay the price. He believed in the wave model for light. His reasoning overlaps Newton's but comes to different conclusions. He was aware of the very great speed of light, but felt that this was incompatible with particles of matter being shot from the source to the eye. Instead, a wave motion propagating through an intervening medium was the more attractive model to him. Huygens was aware also that rays of light could cross one another without disturbing each other. This observation led him to reject particles, which might collide, and made him favor waves that obey a superposition principle. In Huygens' theory, pulses reinforce one another at their common tangent line (Section 6.3). Huygens, unfortunately, did not know about the interference of wave trains and the resulting standing waves, nor about the double-slit interference pattern (Figure 6.21). These phenomena, which strongly confirmed the wave model, were discovered later and would have substantially bolstered Huygens' arguments.

Particle versus wave theory of refraction. Of the two men, Newton had the greater reputation, and his model was accepted by most of his contemporaries. A clear-cut detectable difference between the two models lay in their prediction of the speed of light in dense media like water and glass. According to Newton, the speed was *increased* by the attractive interaction that deflected (refracted) the corpuscles at the surface. According to Huygens, the observed refraction required a decrease in the speed of light (Table 5.1 and Example 7.l). As mentioned above, it was only much later, in the middle 1800s, that Foucault determined that light in fact traveled slower in water than in air. From the modern point of view, both models have weaknesses, but these were only resolved in the twentieth century, after development of the theory

Thomas Young (1773-1829) was an astonishingly versatile figure: physician, linguist, and scientist. While a medical student, Young made original studies of the eye and later developed the first version of the three-primary-color theory of vision. A large inheritance in 1797 enabled him to devote himself primarily to science. After becoming professor at the Royal Institute in 1801, he turned to physical optics and discovered that Newton's work was explainable in terms of waves. Young was also a pioneer in Egyptology and was among the first to try to decipher the Rosetta Stone.

"Those who are attached to the Newtonian Theory of light. .. would do well ... to imagine anything like an explanation of these experiments derived from their own doctrines ..."

Thomas Young
Philosophical Transactions, 1804

To Newton, the downfall of his corpuscular theory would not have been entirely unexpected. Unlike many of his predecessors and followers, Newton did not confuse theory with doctrine: "Tis true, that from my theory I argue the corporeity of light: but I do it without any absolute positiveness ... I knew, that the properties*, which I declared of light, were in some measure capable of being explicated not only by that, but by many other mechanical hypotheses."*

of relativity and quantum mechanics. (A dual theory is now in use, with a wave packet (Section 6.2) playing a central role. Propagation of light is determined by the wave character of the packet, while emission and absorption are determined by the corpuscular aspect of the wave packet as a "chunk" of light.)

Discovery of interference. The next development in the physics of light took place 100 years after the time of Newton and Huygens. Thomas Young conducted experiments on the diffraction of light by two slits; he also observed the patterns we showed for water waves in Fig. 6.2, and he introduced the concepts of superposition and interference of waves in a series of three papers. The two-slit interference experiment was very much more suggestive than the much earlier single-slit diffraction experiments and served to revive the wave model for light. Curiously, Young felt compelled in his first publication to ascribe the original wave model for light more to Newton (who considered it in his treatise) than to Huygens. Young played down Newton's own complete rejection of the wave model.

Young went considerably further in his second and third papers, in which he described the conditions for constructive and destructive interference in terms of the wave path difference measured in "breadths" (wavelengths) of the supposed "undulations," (waves) which differed for different colors. Young was then no longer so respectful of Newton; he used Newton's data to illustrate his own ideas, and he completely rejected Newton's interpretation.

Acceptance of the wave model. The last blow to Newton's corpuscular model was delivered in the middle of the nineteenth century by Foucault's measurements of the speed of light in water. We have already reported that this speed was found to be less than the speed in air, as required by the wave model's explanation of refraction and in contradiction to Newton's prediction based on the corpuscular model. With this achievement, the wave model was unanimously accepted and physicists' attention could turn to new questions: What is the aether? What kind of waves are light waves? How is light emitted and absorbed by matter?

Emission, reflection, and absorption spectra. The emission of light by a hot source means that there is energy transfer from thermal energy of the glowing material to radiant energy of light, which travels to a distant energy receiver. The reverse energy transfer occurs when light is absorbed; then the radiant energy of the light is transferred to a form of energy in the receiver. This is usually thermal energy, but some of it may be chemical energy also (as in photosynthesis, photography, and sunburn).

Emission and absorption by gases. The diffraction grating has been used to decompose the light from many sources into its monochromatic parts. Once the wave nature of light was generally accepted, such studies were used to gain information about the light source itself. It was

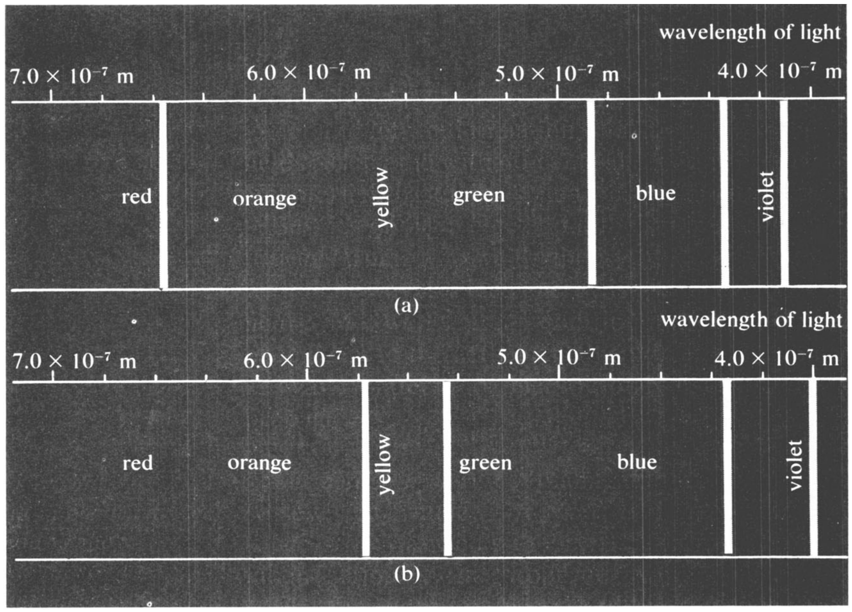


Figure 7.10 Line emission spectra of gases. (a) Hydrogen gas. (b) Mercury gas.

TABLE 7.7 EMISSION SPECTRAL LINES OF SELECTED ELEMENTS

Element (m)	Wavelength (m)
calcium	4.9×10^{-7}
	6.1×10^{-7}
	6.4×10^{-7}
copper	4.6×10^{-7}
helium	5.3×10^{-7}
mercury	4.4×10^{-7}
	5.5 x 10-7
	5.8×10^{-7}
neon	5.4×10^{-7}
	6.4×10^{-7}
	6.5×10^{-7}
sodium	5.9×10^{-7}

found that many glowing gases emitted characteristic line spectra (Fig. 7.10). That is, the spectrum of light they produced did not include all colors, but only certain colors in very narrow bands of wavelengths called spectral lines. These emission line spectra are unique for each element and can be used for identification, just as fingerprints can be used to identify a person (Table 7.7).

Absorption of light by gases, like emission, is selective. That is, gases absorb light only at certain wavelengths or absorption lines. Most of the absorption lines have the same wavelength as emission lines and can be used to identify the presence of a chemical element (Fig. 7.11). Virtually all information about the chemical composition of the sun and stars comes from the analysis of spectral lines. Astronomers have now recorded and analyzed spectra from essentially all of the stars and other objects they have found in the sky; this huge body of evidence leads to a conclusion that may seem disappointing to sci-fi fans: the entire known universe is composed of the same chemical elements as the earth, though in different proportions than found on the earth.

Emission and absorption by solids. Glowing solids, such as light bulb filaments or hot coals, emit continuous spectra (Fig. 7.6). That is, all wavelengths are represented, and not only selected ones as in the

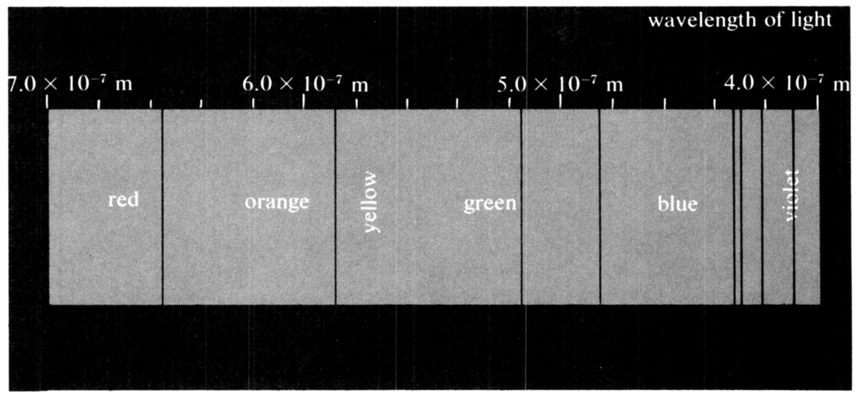


Figure 7.11 Light emitted by the sun. Dark lines (called Fraunhofer lines after their discoverer) are evidence of absorption of light by chemical elements in the gases at the surface of the sun.

line spectrum emitted by a glowing gas. By contrast with line spectra from gases, the continuous spectra from different glowing solid materials are very similar to one another and depend on the temperature but not on the composition of the material.

We have already described the selective reflection and absorption of light that is responsible for the relative brightness and the color of a reflecting surface (Section 5.2). By analyzing the reflected light with a diffraction grating, you can find out whether a color is pure (monochromatic) or mixed. By inference from the colors of the incident and reflected light, you can find which colors (wavelengths) are absorbed.

Micro-domain model for emission and absorption. Are the sources of light macro-domain systems? Since the wavelength of light is at the borderline of the two domains, the sources are in all likelihood very small by macro-domain standards. Another clue comes from the spectra of gases. According to the MIP model, the particles in gases are far apart and do not interact with one another appreciably. Since gases nevertheless emit and absorb light, the source must be a gas particle acting alone. Furthermore, since gases emit line spectra (isolated wavelengths and frequencies), you may conclude that a gas particle acts like a tuned system with several oscillators, one frequency corresponding to each spectral line emitted. In this interpretation of spectral lines, therefore, their frequency is a more significant property than their wavelength.

This model helps to explain why solid materials emit continuous spectra rather than line spectra. In the solid phase, the particles of the MIP model interact strongly with one another. The interaction shifts the frequencies of the individual oscillators so all possible frequencies are represented. Each of these oscillators emits light of its own frequency, but all oscillators together give rise to a continuous spectrum of light.

The frequency of an oscillator responsible for the emission of light

Equation 7.5

speed of light $= v$

wavelength $= \lambda$

frequency $= f$

$$f = \frac{v}{\lambda} = \frac{3\times10^{8}\ m/sec}{5.9\times10^{-7}\ m}$$

$$=5\times10^{14}\ /sec$$

James Clerk *Maxwell* (1831-*1879) was born into a wealthy Scottish family in Edinburgh. After education at Edinburgh and Cambridge, Maxwell was Professor of Physics at Marischal College, Aberdeen, and Kings College, London. He published important papers in 1859-1860 on Saturn's rings and on the kinetic theory of gases. His greatest work, however, was in electromagnetism. Adopting Faraday's theory of fields, Maxwell set out to establish a unified mathematical description of electric and magnetic phenomena. Maxwell's findings, published in 1865 and 1873, are a landmark in theoretical physics.*

can be calculated from the known speed of light and the measured wavelength of one spectral line. To take an example, the frequency of sodium light is found to have a value of enormous magnitude (Eq. 7.5). Since the wavelengths of all visible light do not differ greatly, the frequencies of all the oscillators in the particle model are of similar magnitude. What could be oscillating with such high frequency?

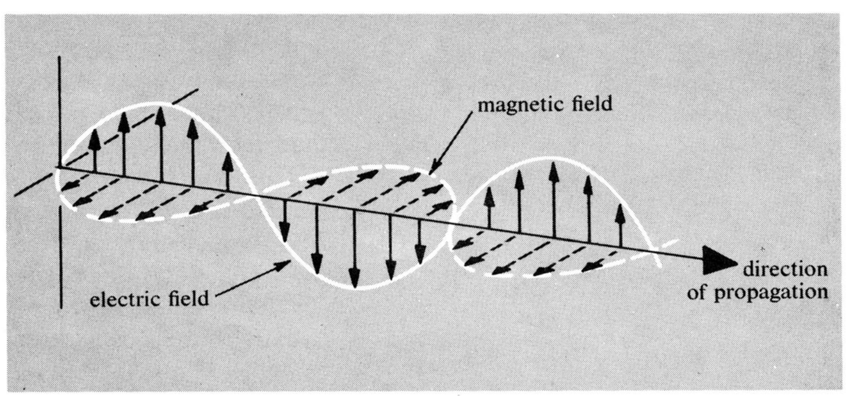
7.3 The electromagnetic theory of light

As we explained in the previous section, the model of light as displacement waves propagating in the aether was generally accepted by the middle of the nineteenth century. In spite of the satisfactory state of affairs, however, there were loose ends yet to be explained. These problems had been pointed out by Newton in his critique of the wave model: no independent evidence of the aether's existence and no mechanism for the emission and absorption of light. Soon, however, there were several developments that led ultimately to a brilliant confirmation of the wave model's applicability to the propagation of light; however, they also required considerable adjustment in the models that scientists had for light waves, and indeed for other physical phenomena.

Maxwell's theory. The first step was a theoretical synthesis, by J. Clerk Maxwell, of the discoveries regarding electric fields, magnetic fields, the magnetic effects of electric currents, and the electric effects of moving magnets, which had been made during the preceding decades. Maxwell came to the conclusion that rapidly vibrating electric charges would generate electric and magnetic fields whose intensity exhibits wavelike patterns, much as a vibrating violin string generates air pressure variations that exhibit wave patterns (Section 7.1).

Maxwell called his waves electromagnetic waves. He viewed them as oscillatory displacements of the aether, again in analogy to sound

Figure 7.12 Electric and magnetic fields in an electromagnetic wave.



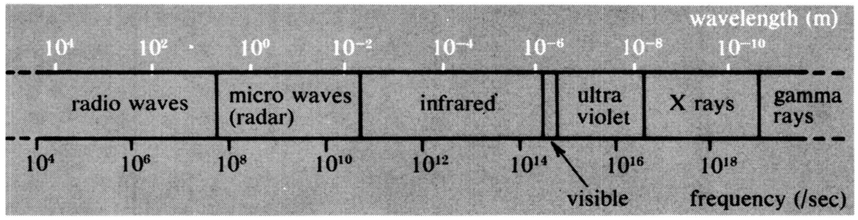


Figure 7.13 The electromagnetic spectrum

waves or water waves. A diagram indicating the electric and magnetic field patterns in such a wave is shown in Fig. 7.12. The electric and magnetic fields are at right angles to each other. At any particular point in space, the intensity of the fields oscillates in magnitude and/or direction with the frequency of the source. The entire pattern shown moves in the direction of propagation with the speed of light.

"The ... difficulties ... which are involved in the assumption of particles acting at a distance ... are such as to prevent me from considering this theory as an ultimate one ... I have therefore preferred to seek an explanation of the facts in another direction. ... The theory I propose may ... be called a theory of the Electromagnetic Field, because it has to do with the space in the neighbourhood of the electric or magnetic bodies."
James Clerk Maxwell
Philosophical Transactions, 1865

Maxwell's calculation indicated that the waves would propagate in aether with a speed of 3×10^8 meters per second. This speed was, to everyone's amazement, just equal to the speed of light measured a few years earlier (Table 7.6). Light waves were therefore identified as electromagnetic waves with a wavelength of about 5×10^{-7} meters. The frequency of these waves is related to their wavelength through the same equation that applies to sound and other waves ($v = f/\lambda$, Eq. 7.5); the frequency of light waves, however, is enormously high compared to that of sound waves. Maxwell's theory demonstrated conclusively that electric charges were the source of the vibrating electric and magnetic fields; thus these charges would have to exist within any object that is a source of light, and in order to generate light, they would have to be able to vibrate at the high frequencies of light waves. Furthermore, if the electric charges vibrate (oscillate) at such a high frequency, they must have an extremely small inertia and be subject to very strong interaction compared to the oscillators that are responsible for generating sound waves. This requirement will turn out to be extremely important in the search for understanding of the structure and constituents of matter.

The characteristic spectra of gases exhibit sharp frequencies and are evidence that gases contain "tuned" electrically charged systems capable of oscillating at well-defined high frequencies. In Chapter 8 we will explore the sources of light waves, which are now identified with the atoms and molecules composing all matter.

The electromagnetic spectrum. Maxwell's theory suggested strongly that electromagnetic waves should exist with frequencies different from those of light. One only had to arrange for electric charges to vibrate at a lower frequency. Heinrich Hertz (1857-1894) succeeded in generating waves with a wavelength of a few centimeters (rather than the 5×10^{-7} m wavelength of light) by making sparks in simple electric circuits, and Marconi (1874-1937) turned this discovery to practical use in his invention of the wireless telegraph (radio). Many other forms of electromagnetic (E-M) radiation have since been discovered. Such E-M waves are extremely useful; for example, radio and radar waves can transmit information over great distances, and x-rays and infrared can create images of objects that are otherwise invisible. The entire frequency range is called the electromagnetic spectrum (Fig. 7.13). It includes not only visible light and radio waves, but also X-rays, ultraviolet and infrared radiation, radar, and even the electric and magnetic fields associated with 60-cycle alternating house current (Section 12.4). Visible light actually spans only a minute portion of the spectrum.

"It appears therefore that certain phenomena in electricity and magnetism lead to the same conclusion as those of optics, namely, that there is an aethereal medium pervading all bodies, and modified only in degree by their presence ..."
James Clerk Maxwell
Philosophical Transactions, 1865

Besides giving a clue about the sources of light and vastly expanding the spectrum, Maxwell's theory made it possible to associate energy with light waves, since energy is associated with electric and magnetic fields. Maxwell's great contribution, however, also called attention once again to the aether, the medium in which light waves, now viewed as wave patterns of electric and magnetic fields, were believed to propagate.

The aether mystery. You might think that the existence of the aether, so severely criticized by Newton, was now firmly established with the success of Maxwell's theory. Far from it! Now that aether had to be taken seriously, its properties were investigated more thoroughly.

The first question was asked by Maxwell himself: What about the motion of the aether? Clearly, he reasoned, light waves traveling at a certain speed relative to the aether would be observed to travel at a different speed relative to objects moving with respect to the aether. Nobody knew, of course, what objects moved with respect to the aether, but the planets' relative orbital motion at different rates made it impossible for the aether to be at rest with respect to all of them at the same time.

Albert Einstein (1879-1955), perhaps the greatest theoretical physicist since Isaac Newton, was born in Ulm, Germany and educated in Munich and Switzerland. After graduation he could not obtain a university teaching position and had to accept the obscurity of a minor post at the Berne Patent Office. This obscurity ended dramatically in 1905 when Einstein published five important papers, including two that shook the scientific world to its foundations – one on the photoelectric effect, and the other on the special theory of relativity. In 1933, Einstein resigned as Director of the Kaiser Wilhelm Institute of Physics in Berlin as a protest against Hitler's fascist policies. He emigrated to the United States, where he spent the rest of his life working at the Institute for Advanced Study in Princeton.

Many experiments were carried out to detect motion of the earth relative to the aether by comparing the speed of light measured under many different conditions: parallel to the earth's orbital motion around the sun, perpendicular to this motion, inside rapidly moving liquids, light generated on earth and coming from moving sources, and so on. The results were negative; the speed of light gave no evidence that the earth moved relative to the aether. Was the earth, then, really at rest in the aether, while the entire universe moved around the earth? More than 300 years after Copernicus this was not an acceptable hypothesis.

There were other mysteries about the aether; we pointed out above, in Section 7.2, that the behavior of light and sound was not consistent: light traveled slower in denser materials (water and glass) than air but sound traveled faster in such materials. This apparent contradiction raised doubts about the similarities between the medium for sound and the aether. Such concerns became more serious when scientists started actively investigating the properties of the aether, using Maxwell's theory of electromagnetism. For sound (and other waves), the faster the speed of the wave, the lower the inertia and the higher the interaction among the oscillators of the medium. However, applying this kind of thinking to light and its extremely high speed meant that the aether would have to be made up of oscillators with contradictory properties: an inertia that was much, much lower, yet a strength of interaction that was much, much stronger, than in any existing substance!

The theory of relativity. Attempts to resolve these contradictions made little progress until Albert Einstein, early in the twentieth century, approached the subject from a new, apparently unrelated, direction; Einstein seriously pursued the problem of how the interaction of electrically charged bodies and the light waves they generate would appear to two observers in relative motion. In particular, how would a light wave in vacuum appear to an observer who moves alongside it, with the same speed? Maxwell's theory excluded the possibility of a stationary light pulse.

Einstein took the viewpoint that the interaction of two electrically charged objects should depend only on their motion relative to one another and not on their common motion relative to some outside reference frame. To reconcile this requirement with the known laws governing electric and magnetic interactions, Einstein found it necessary to abandon the aether and to modify the commonsense concepts of space and time.

At the basis of Einstein's reasoning was an operational approach to the methods by which observers in relative motion can communicate their

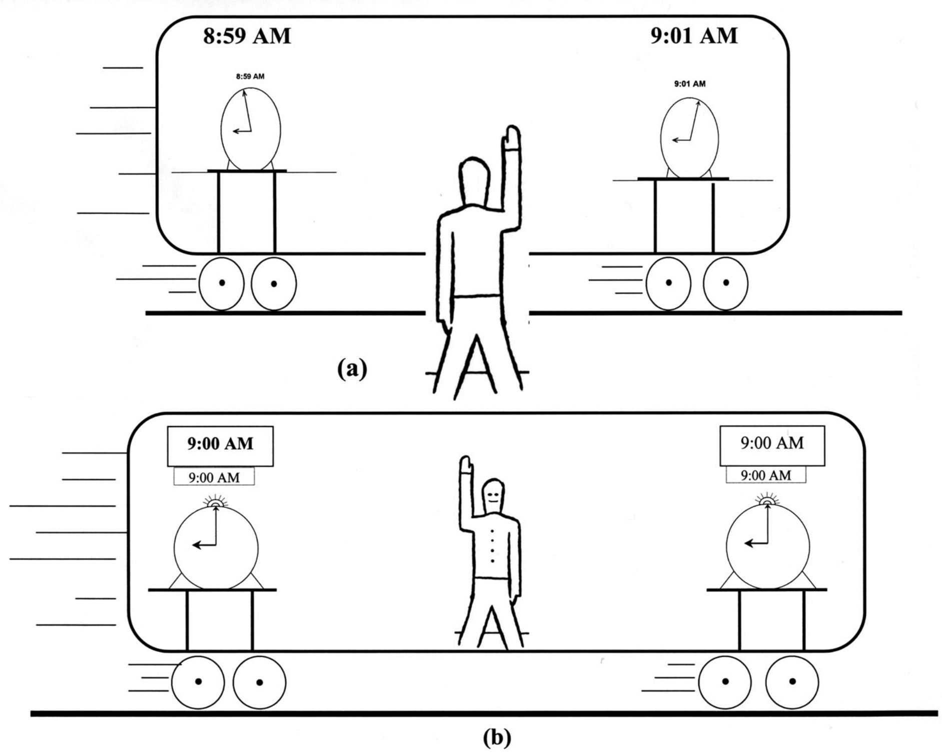


Figure 7.14 Two clocks on a train, as observed by two observers in relative motion. The effect is vastly exaggerated and would not be observable in a train.
(a) To the observer opposite the car's center, but off the train, the clock at the rear of the train would appear to be running two minutes behind the clock at the front of the train.
(b) To the observer at the car's center on the train, the two clocks would appear perfectly synchronized.
This paradoxical behavior is the direct result of the assumption that the speed of light is the same for all observers, which also seems to contradict our experience but is extremely well documented. It would seem difficult to build a reliable theory of physics based on such anti-intuitive ideas, but Einstein did exactly that. His theory of relativity provides a solid basis of understanding for all situations in which velocities near the velocity of light are involved.

Einstein's disruption of the old Newtonian scheme inspired J. C. Squire to add to Alexander Pope's couplet:

"Nature and Nature's Laws
lay hid in night.
God said, Let Newton be,
and all was light."

one of his own:

"It did not last: the Devil
howling, `Ho,
Let Einstein be,' restored the
status quo."

observations to one another. Instead of assuming that this could be done, as had Galileo and Newton, he described how observers must use light signals (the fastest known method of communication over a distance) to standardize their instruments for measuring distances, time intervals, and so on. Built into the scheme were the experimental results, which indicated that the observed speed of light was not influenced by motion of neither the light source nor the light detector.

Einstein's results, embodied in his theory of relativity, have substantially influenced both science and philosophy. Einstein demonstrated that several of our apparently "intuitive" ideas about the physical world had to be discarded. The theory of relativity was both consistent and comprehensive, and it required us to accept several new ways of thinking that seemed to conflict with intuition: First, two events that are simultaneous for one observer are not simultaneous for a second observer in motion relative to the first (Fig. 7.14). Second, if observer A is moving with respect to observer B and they both measure the length of a given object and the duration of a given event, (using standard rulers and clocks moving with them), they will *not* find the same results. Third, an automobile moving at exactly 75 miles per hour passing another at exactly 55 miles per hour is not traveling at exactly 20 miles per hour relative to the second car (Section 2.2), though the difference is insignificant for such slowly moving objects. Finally, the answer to Einstein's original question about the observer "catching up" with the light wave in vacuum is deceptively simple: the properties of space and time are such that this can never happen!

Summary

Sound and light are intermediaries in interaction-at-a-distance between a source and a receiver. The modern understanding of both phenomena is achieved with the help of a wave model.

Sound waves are pressure waves in solids, liquids, or gases. Associated with the pressure variations are displacements of micro-domain particles making up the material. The tone or pitch of sound is associated with the wave frequency, the intensity with the wave amplitude. The speed of sound in air is 344 meters per second.

The human ear can detect sound waves whose frequencies fall between about 20 vibrations per second and 20,000 vibrations per second, with most speech using waves from 200 to 2000 per second. The wavelength of sound waves in speech, therefore, is comparable in size to the human body and to many environmental objects in the macro domain. These sound waves are so strongly diffracted by objects of this size that no sharp acoustic images can be formed.

Light waves are electromagnetic waves that propagate in vacuum and in various media. Visible light is only a very narrow portion of the electromagnetic spectrum, which also includes radio, microwaves (radar), infrared, ultraviolet, X-ray, and nuclear radiation. The color of light is associated with the wave frequency, the intensity with the wave amplitude. The speed of electromagnetic waves in vacuum is 3×10^8 meters per second.

The wavelength of visible light in air is very small, about 5 x 10^{-7} meters. Evidence of the wave nature of light is therefore difficult to obtain through experiments in the macro domain. The ray model describes light very well until it interacts with objects at the lower limit of the macro domain. Then the effects of diffraction and interference can be observed and the inadequacies of the ray model are revealed.

As a matter of fact, light spans the micro, macro, and cosmic domains. The speed of light is so great that it traverses the cosmic distance from the earth to the moon in about 1 second. The wavelength of light is in the micro domain. And the light rays (pencils of light) that make possible human vision are in the macro domain. No wonder that so much controversy surrounded the models for light! For this very reason, however, light has been a powerful tool in the study of systems in the micro and cosmic domains.

One of the most revolutionary consequences of the electromagnetic theory, as applied by Einstein, was to eliminate the need for the existence of the aether. Light waves propagate in vacuum, without a medium. There is no medium, therefore, that could serve as a special reference frame for measurements of the speed of light. The startling consequences of this conclusion have changed out view of space and time.

List of new terms

octave	shock waves	infrared
ultrasonics	spectrum	double refraction
sonar	absorption	line spectrum
sub-audible waves	emission	electromagnetic waves
seismic waves	ultraviolet	aether
theory of relativity		

List of symbols

v	wave speed	Δs	distance traversed
λ	wavelength	Δt	time interval
f	frequency	d	distance between grating slits
c	speed of light in vacuum	l, L	distances in experimental arrangements
θ	angle		
n	index of refraction		

Problems

1. Prepare several more or less well-tuned systems that act as sound sources (rubber band, stretched wire, air in a bottle, glass of water) and experiment to produce musical notes with them.
   (a) Describe how you can change the pitch of the note (i.e., tune the system) in terms of the wave model for sound.
   (b) Relate the pitch of the sound to the wave speed in the medium and the dimensions of the tuned system.
2. Mersenne's shorter brass wire was 22 centimeters long (Section 7.1). Estimate the wave speed on the wire in his experiment. (Note: the

speed of a mechanical wave on a wire is usually *not* equal to the speed of a sound wave in the wire material. Mechanical waves of the wire as a whole involve macro-domain displacements; whereas sound waves in the material, set up, for example, by scraping the wire with a file, involve micro-domain displacements of the many interacting particles making up the wire.)

3. Explain the following three statements with the help of the gasbag model for sound waves and the conditions for wave propagation. If necessary, make hypotheses about the materials to explain the observations. (Note: a successful explanation is evidence that the hypothesis is valid.)
(a) The speed of sound in air at room temperature does not change if the air pressure is increased or decreased.
(b) The speed of sound in air increases with the temperature. At the boiling temperature of water it is 386 meters per second.
(c) The speed of sound in carbon dioxide is less, and in hydrogen gas is more, than in air at the same temperature.

4. Complete the wavelength column in Table 7.2.

5. Calculate the frequencies and wavelengths of the highest and lowest notes on an 88-key piano.

6. An organ pipe that plays a B note is filled with hydrogen gas. What note will it play then?

7. Look for "sound shadows" produced by buildings or other large obstacles. Describe the relative position of the sound, source, obstacles, and receiver for a significant shadowing to be observable. (Note: you may use your ear as receiver, but you will need a reliable and cooperative sound source, such as a friend with a musical instrument.)

8. Explain the operation of sonar with respect to the following. (a) What size underwater objects can be reliably detected by sonar of 20,000 vibrations per second? Justify your estimate. (b) The echo from an underwater object is received 5 seconds after the sonar pulse was emitted. How distant is the object?

9. Find the wavelength of bat "sonar" pulses. (frequency about 10^5 /sec)

10. (a) Estimate the wavelength of seismic waves.

(b) Would you recommend that geologists use a ray model or a wave model for seismic waves? Explain your answer and any limitations it may have.

11. Use a fine, regularly woven fabric to measure the wavelength of light approximately (Figs. 7.2, 7.3, Problem 5.9).

12. Look at the glancing reflection of a bright, medium-distant light source in a phonograph record, compact disc (CD) or digital videodisc (DVD). (See figure at left.)
(a) Describe your observations and explain them in terms of the wave theory of light.
(b) Make appropriate measurements and calculate the approximate spacing of the tracks on the record, CD, or DVD.

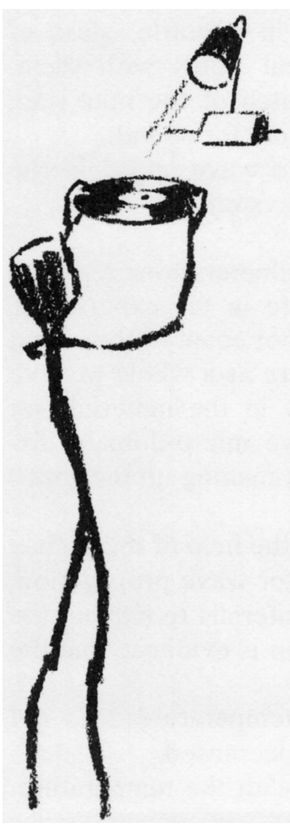


Diagram for Problem 12

Problems 13-16. To observe interference of light, you will need a narrow light source and one or more very thin slits. The simple slit holder shown in Fig. 7.15 below gives satisfactory results. A "source slit" aimed at a lamp acts as a narrow light source. A diffraction grating, a single slit, or a double slit can be attached to the opening in near your eye so as to view the light from the source slit.

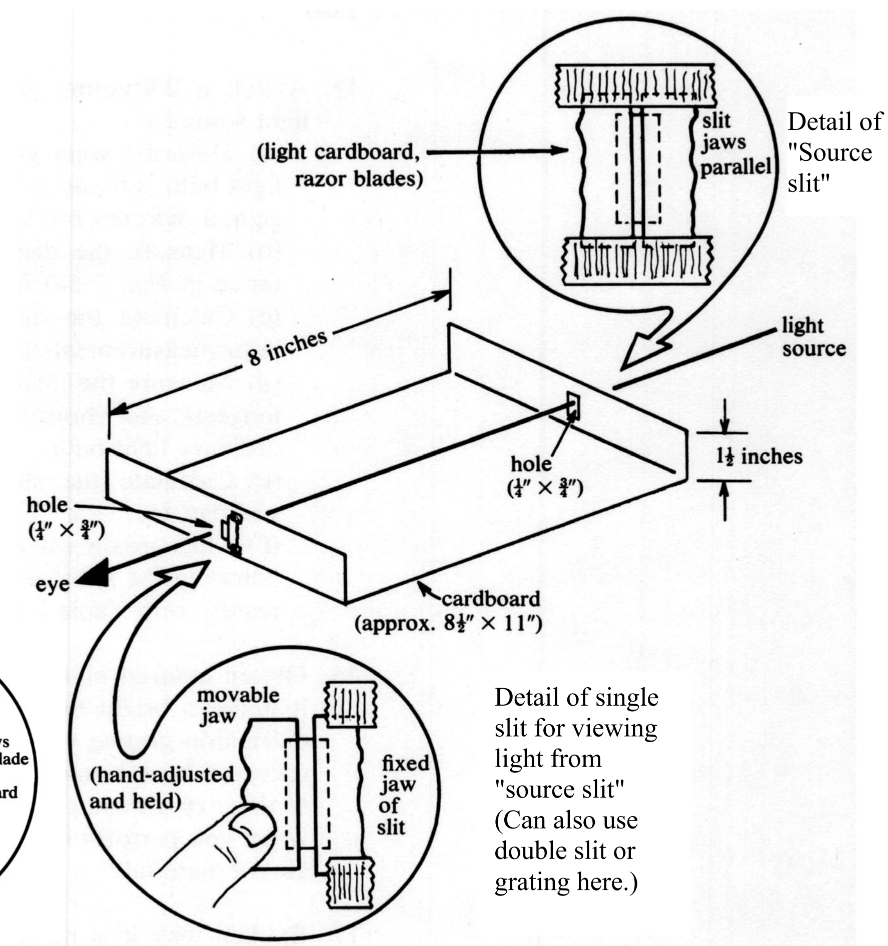


Figure 7.15 (to right) Slit holder and slits for Problems 13-16. The slits shown in the detailed drawings above and below the slit holder may be attached to the slit holder or may be built on separate pieces of cardboard that are attached to the slit holder with paper clips.

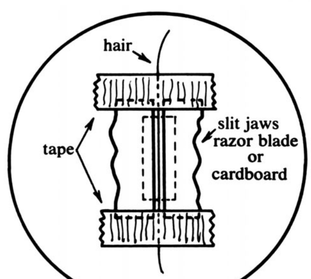


Figure 7.16 (above) Construction of a double slit. The slits should be as narrow as possible, with a hair centered in a cardboard slit (Fig. 7.15) to make it into a double slit. Keep the slit jaws parallel. You may need to work with a magnifier

13. Make a variable-width slit (Fig. 7.15) and attach it to the slit holder. Look at the light bulb through the holder and observe the single-slit diffraction pattern. Describe the colors, the number of bright fringes, and the spacing of the fringes while you vary the slit width.

14. Make a double slit (Fig. 7.16) and attach it to the slit holder. Look at a light bulb through the holder and observe the double-slit diffraction pattern.
    (a) Describe your observations (color, number of bright fringes, position of dark fringes).
    (b) Measure the position of the dark fringes and the slit dimensions as well as you can. Use these to calculate (roughly) the wavelength of light (Section 6.4, Problem 6.12).

15. Attach a plastic diffraction grating to the holder and look at various light sources.
(a) Describe what you observe when you look at an ordinary light bulb, a fluorescent bulb, a red "neon" sign, a green "neon" sign, a mercury arc lamp and a sodium vapor lamp. Sodium vapor lamps are often used in street lights because of their high efficiency; their light is very strongly yellow and seems harsh.
(b) Measure the displacement of the diffraction images (distance labeled "l" in Fig. 7.5a) for the spectral lines of one element.
(c) Calculate the distance between slits in the grating from your measurements using Eq. 7.3 and the known wavelength of the light from that element (see Table 7.7).
(d) Measure the displacement of the diffraction images for the longest- and shortest-wavelength light you can see from an ordinary light bulb.
(e) Calculate the shortest and longest wavelengths of light you can see.
(f) Measure the wavelengths of the various colors as you see them in the spectrum of an ordinary light bulb. Compare your results with Table 7.5.

16. Obtain colored cloth or paper (preferably not glossy) and place it under a bright lamp. Look at this colored material through a diffraction grating on your slit holder (Fig. 7.15). Describe the light that is reflected by the material. (You may wish to compare the result of this experiment carried out in bright sunlight with that obtained when you perform it under artificial light.) Why should you avoid glossy material?

17. Explain why it is more accurate to say that the color of light is associated with its frequency rather than with its wavelength.

18. Describe your evaluation of the disagreement between Huygens and Newton. Optional: Do additional reading on the subject.

19. (a) Calculate the wavelength range of the standard AM broadcast band (frequency 5.6 –1600 kHz, 1 kHz = 1,000 /sec)
(b) Calculate the wavelength range of FM broadcasts. (frequency 88–106 MHz, 1 MHz = 1,000,000 /sec)
(c) Explain why hilly terrain interferes with FM broadcast reception much more seriously than with AM radio reception.

20. Identify one or more explanations or discussions in this chapter that you find inadequate. Describe the general reasons for your dissatisfaction (conclusions contradict your ideas, or steps in the reasoning have been omitted; words or phrases are meaningless; equations are hard to follow; and so on) and pinpoint your criticism as well as you can.

Bibliography

L. Barnett, *The Universe and Dr. Einstein*, Harper and Row, New York, 1948.

A. H. Benade, *Horns, Strings, and Harmony*, Doubleday, Garden City, New York, 1960.

R. E. Berg and Stork, D. G., *The Physics of Sound*, 2nd Ed., Prentice Hall, 1995.

A. Einstein, *Relativity*, Doubleday, Garden City, New York, 1947.

M. Gardner, *Relativity for the Millions*, Macmillan, New York, 1962.

D. E. Hall, *Musical Acoustics*, 2nd Ed, Brooks/Cole Publishing, 1991.

C. M. Hutchins and F. L. Fielding, "Acoustical Measurement of Violins," *Physics Today* (July 1968).

C. Huygens, *Treatise on Light*, Dover Publications, New York, 1912.

W. E. Kock, *Sound Waves and Light Waves*, Doubleday (Anchor), Garden City, New York, 1965.

E. Mach, *The Principles of Physical Optics: A Historical and Philosophical Treatment*, Dover Publications, New York, 1952.

G. Murchie, *Music of the Spheres*, Houghton-Mifflin, Cambridge, Massachusetts. See Chapters 11 and 12.

J. Rigden*, Physics and the Sound of Music*, Wiley, 1977.

T. D. Rossing, *The Science of Sound,* 2nd Ed, Addison-Wesley 1990.

B. Russell, *The ABC of Relativity*, New American Library, New York, 1959.

W. A. M. Bergeijk and others, *Waves and the Ear*, Doubleday, Garden City, New York, 1960.

E. Wood, *Experiments with Crystals and Light*, Bell Telephone Laboratories, Murray Hill, New Jersey, 1964.

Articles from Scientific American. Some or all of these, plus many others, can be obtained on the Internet at http://www.sciamarchive.org/.

E. D. Blackham, "The Physics of the Piano" (December 1965).

E. Bullard, "The Detection of Underground Explosions" (July 1966).

E. E. Helm, "The Vibrating String of the Pythagoreans" (December 1967).

E. N. Leith and J. Upatnieks, "Photography by Laser" (June 1965).

The entire September 1968 issue of *Scientific American* is devoted to "light." The following articles are of particular interest:

P. Connes, "How Light is Analyzed."

G. Feinberg, "Light."

D. R. Herriott, "Applications of Laser Light."

U. Neisser, "The Processes of Vision."

A. L. Schawlow, "Laser Light."

V. F. Weisskopf, "How Light Interacts with Matter."

In the previous chapters we have had frequent occasion to contrast direct observations of nature with the models scientists have made for interrelating these observations. The many-interacting-particles (MIP) model for matter has been a powerful tool to this end. In this model, matter is composed of micro-domain particles that are in motion and that interact-at-a-distance with one another by means of an intermediary field. During the nineteenth century the various types of energy—kinetic, thermal, elastic, chemical, phase, electromagnetic field (including radiant), and gravitational field—were investigated intensively. As we pointed out in Section 4.5, and as we will describe in greater detail in later chapters, it is possible to explain thermal, elastic, chemical, and phase energies in terms of the kinetic energies of all the particles and the field energy arising from their interaction.

During the twentieth century, physicists have set themselves the goal of explaining all macro-domain phenomena, including perhaps even life, in terms of these particles, which are called atoms or molecules. Great strides in formulating these explanations have been made, except in the case of phenomena involving the gravitational field. Even Einstein, who reformulated the gravitational interaction in a very novel and general way in his general theory of relativity, was not able to relate gravitation effectively to the electric, magnetic and other fields.

One of the most notable areas of progress has been in the invention of models for atoms themselves. Instead of being conceived of as simple point-like particles with no internal structure, atoms are now viewed as complex systems composed of simpler constituents. The properties of atoms are explained in terms of the arrangement and motion of the constituents. In this chapter we will review some of the studies of the current century that have led to the presently accepted models for atoms.

8.1 The electrical nature of matter

Dalton's atomic theory. When John Dalton proposed his atomic theory of chemical reactions, the particles were called *atoms* (from the Greek word for "indivisible") because they were conceived of as being ultimate constituents that would permit no further subdivision. In this view, which gradually became accepted during the first half of the nineteenth century, each chemical element is composed of a different kind of atom. There were about 90 kinds of atoms, each with its own characteristics (Fig. 8.1). The atoms were believed to have properties that could account for the chemical activity and various other macro-domain properties (hardness, appearance, melting and boiling temperatures, and so on) of specimens of the element.

"... the existence of these ultimate particles of matter can scarcely be doubted, though they are probably much too small ever to be exhibited by microscopic improvement. I have chosen the word atom to signify these ultimate particles...."
John Dalton
A New System of Chemical Philosophy, 1808

Electric conduction in solid and liquid materials. As soon as the existence of Dalton's atoms became non-controversial, questions arose about the "intrinsic" properties of the atoms and whether models could be constructed to account for them. In other words, scientists asked in what way an oxygen atom differed from a nitrogen atom, why solid copper conducts an electric current while solid sulfur does not, and how

Humphry Davy (1778-1829) was director of an institution in Bristol, England, that investigated the medicinal properties of various gases when he published the work on nitrous oxide (1800) that made him famous. He was appointed to the Royal Institution in London, where he continued his research. Davy is remembered as the inventor of the coal miner's safety lamp, and he was the first to explain electrolysis and to suggest that electricity and chemical interactions are due to the same ultimate cause (the electric field, as we now call it)

Michael Faraday (1791-1867), a blacksmith's son, was apprenticed in 1804 to learn the bookbinding trade. The boy used his spare time to read books on electricity and chemistry, and in 1812 he attended a series of lectures by Humphry Davy. Faraday wrote a summary of the lectures and sent it to Davy, along with a request for a job. The summary must have been brilliant; it attracted Davy's attention, and he appointed Faraday as a laboratory assistant at the Royal Institution. Faraday eventually succeeded Davy as Director, and Davy was to boast that of all his discoveries the greatest was Michael Faraday. Many historians rank Faraday as the greatest of all experimental physicists. Although he knew no mathematics, he invented the field concept and formulated a complete descriptive theory of electricity.

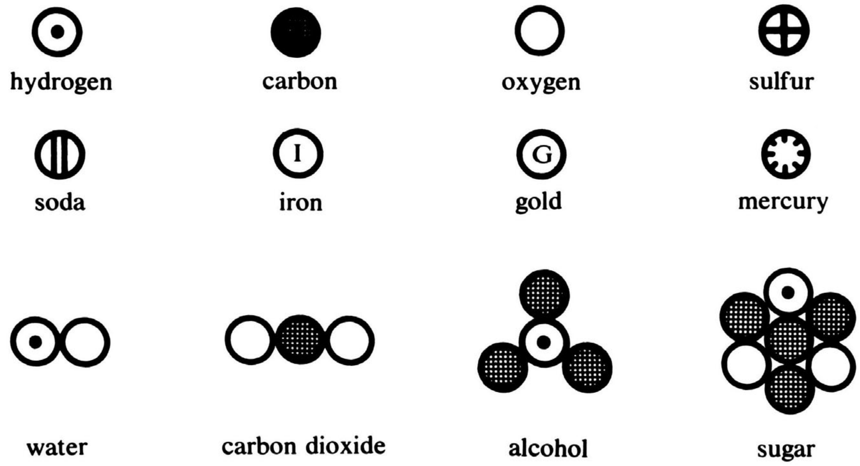


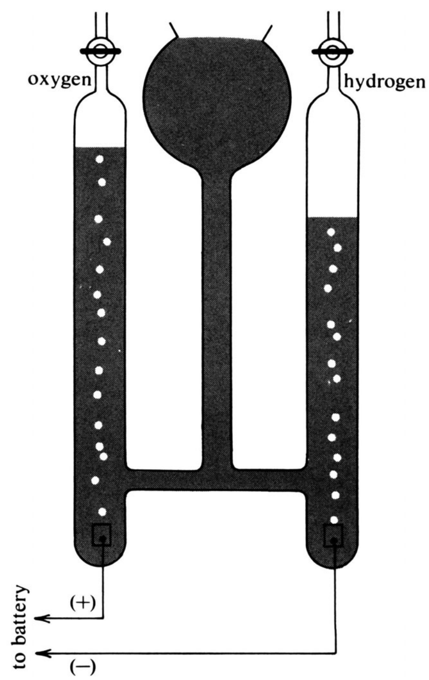
Figure 8.1 Dalton's symbols for the atoms of common elements. The bottom row shows his models for the more complex "atoms" of common compounds.

the metal sodium and the gas chlorine could interact to produce the white crystalline substance sodium chloride (ordinary table salt). Furthermore, a solution of sodium chloride and water conducts an electric current, but solid sodium chloride and pure water separately do not.

Electrolysis. Sir Humphry Davy and Michael Faraday found another effect: melted table salt (sodium chloride) can be decomposed into sodium and chlorine by the passage of an electric current, but the element copper is not modified on the macro level by an electric current. The decomposition of sodium chloride is an example of *electrolysis* (using electricity to divide something into its parts). An electric current can

Figure 8.2 (to the right)
The electrolysis of water to form hydrogen and oxygen gases. Note the ratio of the volumes of gases produced. What do you conclude about the composition of water? On this basis, can you suggest a modification of Dalton's symbol for water above in Figure 8.1?

It is interesting that Dalton knew about this experiment, and others giving similar results about the simple whole number ratios of the volumes of combining gases, but he did not accept these findings, possibly because the particular model of atoms he had developed couldn't explain it!



OPERATIONAL DEFINITION The quantity of electric charge is measured by the mass of hydrogen it liberates in the electrolysis of water. The unit of electric charge, called the Faraday *(symbol F), is the quantity of electricity associated with 1 gram of hydrogen.*

Sir Joseph John Thomson (1856-1940) was born in Manchester and studied at Cambridge, England. In 1884 he was appointed to what was then the most prestigious position in physics, the Cavendish chair of physics at Cambridge. His discovery of the electron and its properties in 1897 won him a Nobel Prize in 1906.

"I hope I may be allowed to record some theoretical speculations ... I put them forward only as working hypotheses, ... to be retained as long as they are of assistance ... The phenomena in these exhausted tubes reveal to physical science a new world—a world where matter may exist in a fourth state, where the corpuscular theory of light may be true, and where light does not always move in straight lines, but where we can never enter, and with which we must be content to observe and experiment from the outside."
Sir William Crookes
Philosophical Transact., 1879

also break up (electrolyse) water, producing hydrogen gas and oxygen gas in a volume ratio of two to one (Fig. 8.2). Faraday concluded from these observations that atoms of matter must be endowed with electrical charges. In fact, the mass of material produced in electrolysis can be used to formulate an operational definition of the quantity of electric charge; the details of the definition are in the left margin.

Franklin's electric fluid. Even air permits the passage of an electric current, as in lightning or an electric spark. Certain materials can be given an electric charge by rubbing. Benjamin Franklin studied these phenomena and concluded, as explained in Section 3.5, that there was one electric fluid whose presence in greater or lesser amounts showed up as positive and negative charges. According to Faraday, these electric charges had to originate within the atom. The generation of electromagnetic waves in association with sparks (Section 7.3) was further evidence that the constituents of matter had electrical properties. Arrayed against these conclusions was the lack of electric effects in ordinary pieces of matter such as a glass of water, a coin, or the air we breathe.

Electric conduction in air. William Crookes (1832-1919) and J. J. Thomson studied in detail the conduction of electric current through air in a glass tube. Fig. 8.3 shows their apparatus. Electric terminals were connected to metal pieces (called electrodes), which were sealed through the ends of the tube. The gas in the tube could be pumped out so as to produce a partial vacuum inside. At ordinary atmospheric pressure, sparks jumped from one electrode to the other. At low pressure, the gas became luminous. But at very low pressure, the glass of the tube itself glowed—evidence of interaction-at-a-distance between the electrodes and the glass. When a metal screen with a fluorescent covering was placed in the tube, it glowed and cast a "shadow" on the glass tube, presumably by intercepting the "rays" that were passing between the electrodes and the glass. In this way it was possible to show that the negative electrode, called the *cathode,* was the source of the rays, which were called cathode rays, and the apparatus became known as a "cathode ray tube." Whether the rays were a stream of particles or a wave phenomenon was not yet known.

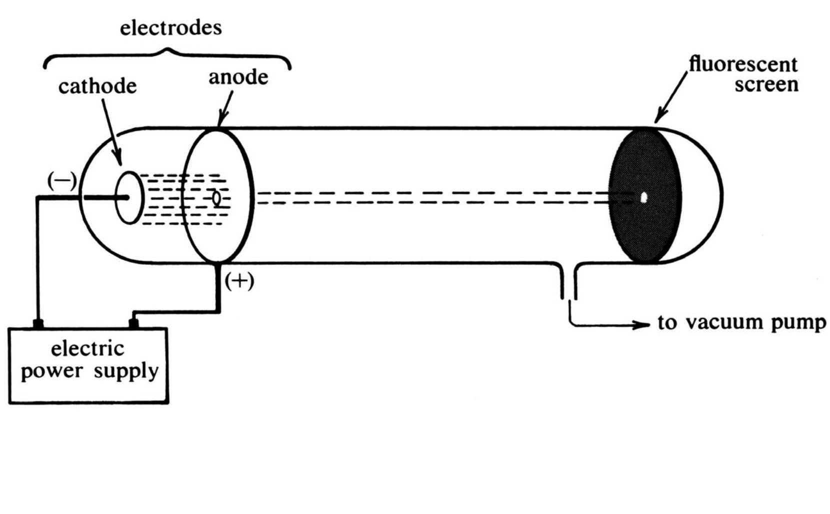


Figure 8.3 (to right) An electric current can pass through low-pressure air in a glass tube, giving rise to many interesting luminous effects. This simple "cathode ray tube" stimulated many fruitful investigations, giving birth to the electron, electronics, the TV tube, X-rays, and other discoveries.

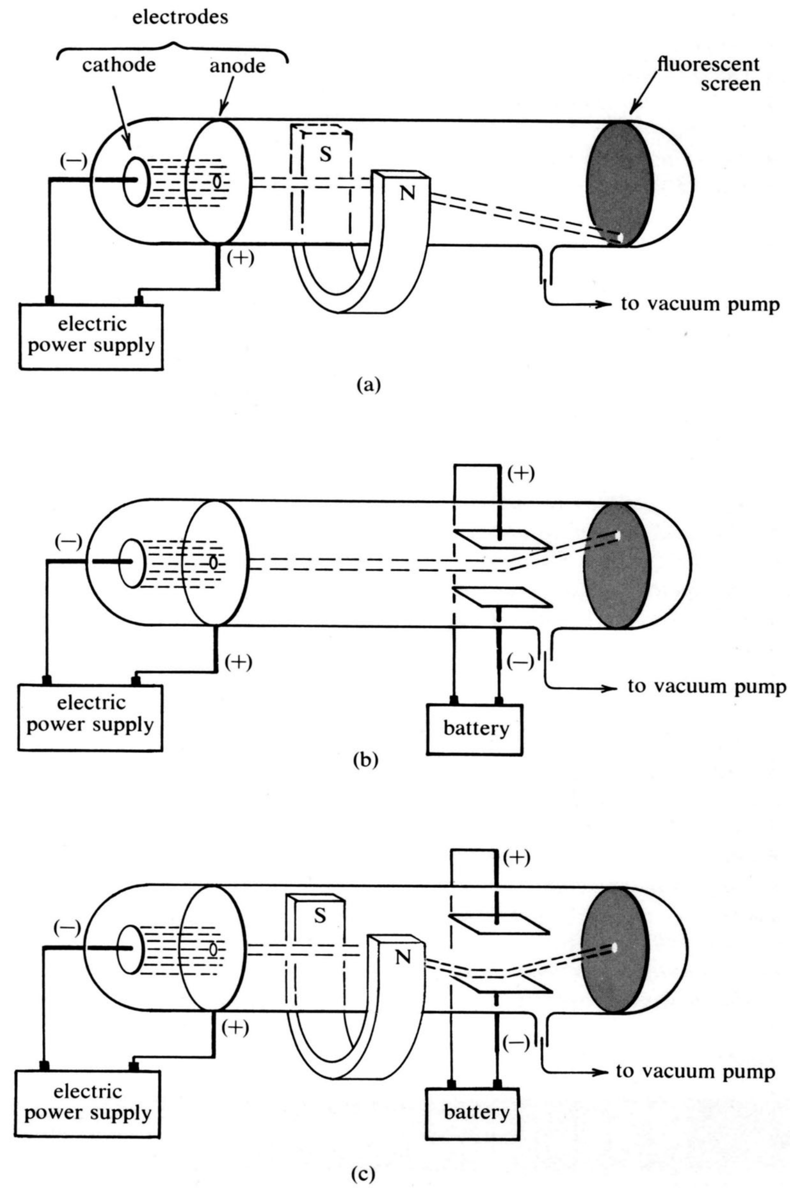


Figure 8.4 (to right). Deflection of cathode rays by magnetic and electric fields. (a) Deflection by a magnetic field. (b) Deflection by an electric field. (c) Deflection by both magnetic and electric fields.

"... The most diverse opinions are held as to [cathode] rays; according to the almost unanimous opinion of German physicists they are due to some [wave] process in the aether to which—inasmuch as in a uniform magnetic field their course is circular and not rectilinear—no phenomenon hitherto observed is analogous; another view of these rays is that, so far from being wholly aetherial [wave-like], they are in fact wholly material, and that they mark the paths of particles of matter charged with negative electricity . . . I can see no escape from the conclusion that they are charges of negative electricity carried by particles of matter. The question next arises, What are these particles? are they atoms, or molecules, or matter in a still finer state of subdivision?"

J. J. Thomson
Philosophical Magazine, 1897

Electrons. By collecting the rays, letting them interact with electrically charged bodies, deflecting them with magnets (Fig. 8.4), and quantitatively measuring the effects, J. J. Thomson was able to show that the cathode rays carried electric charge and inertial mass in a fixed ratio. He therefore proposed a particle model to explain the interaction-at-a-distance between the cathode and glass: many tiny, identical micro-domain particles with a definite mass and negative charge are emitted by the cathode and acquire kinetic energy on being attracted by the positive electrode (the *anode).* Thomson adopted the term *electron* for these particles. Robert Millikan later measured the electric charge of a single electron in 1909 by means of his ingenious oil drop experiment, the first measurement of a physical

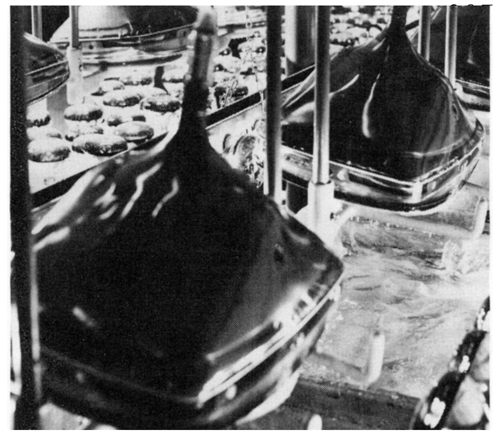


Figure 8.5 Color television tubes in an assembly line are being baked over heat lamps to bond a special phosphorescent material to the inside of the glass screen. TV tubes are essentially cathode ray tubes; the electrons are ejected by the cathode (at top of photograph), formed into a beam, and allowed to strike the screen (at bottom), transferring some of their energy to the phosphorescent material, which emits light of specific colors. Magnetic fields are used to direct the electron beam to specific points on the screen and thus form the image we see.

Robert A. Millikan (1868-1953), a physicist at the University of Chicago and an experimentalist of extraordinary talent and patience, developed his "oil drop experiment" by which changes of one electron charge could be detected. He was the first to carry out a direct measurement of the charge of the electron. He also verified Einstein's hypothesis of light quanta. His painstaking work on the charge of the electron and the photoelectric effect won him the Nobel Prize in 1923.

Millikan found that the charge of one electron was 1.7×10^{-24} *faraday. Hence 1 gram of hydrogen has* 6.0×10^{23} *electrons associated with it.*

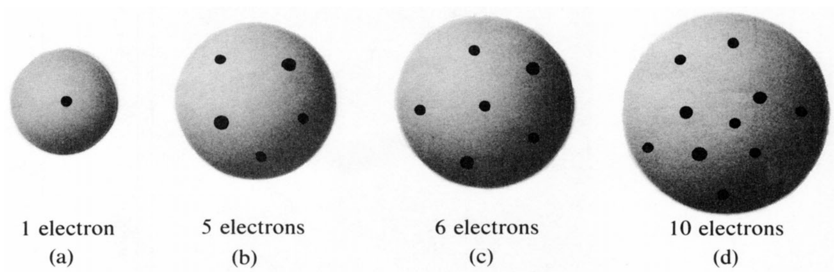
quantity in the micro domain. Millikan's value of the electric charge plus Thomson's value for the charge-to-mass ratio yielded the actual mass of the electron. Modern application of the cathode ray tube has created the electronics industry and revolutionized communications (Fig. 8.5).

Electron waves. According to J. J. Thomson in 1897, cathode rays were minute, electrically-charged "corpuscles." However, physicists in Germany, where Crookes had originally discovered cathode rays, thought of the cathode rays as similar to light and other types of electromagnetic waves (including the recently discovered X-rays), and they attempted to explain the cathode rays on the basis of a wave motion of the "aether" (see quote from Thomson on previous page). Thomson overcame these arguments by the turn of the century with a masterful series of experiments, volumes of data and many outstanding papers demonstrating how his data confirmed the predictions of the particle model and conflicted with those of the wave model.

As a matter of fact, no one at the turn of the century actually tested the cathode rays, directly, for wave properties. One reason for this was Thomson's experiments and tightly-woven arguments; another was the fact that electric charges had always been associated with matter, and no one had ever observed charges transported from place to place by waves. In any case, no one pursued this for almost 30 years, until, as we explain below in Section 8.4, a wave model for all matter was proposed on theoretical grounds. As a result, in 1926, Davisson and Germer performed an experiment in which cathode rays were intercepted by a nickel crystal, which had regularly arranged atoms that functioned like a diffraction grating. Davisson and Germer found a diffraction pattern, as predicted by a wave model for the cathode rays! However, the wavelength of the electrons was extremely short, much shorter than the wavelength of visible light.

The present view, therefore, is that the propagation of electrons is best described as a wave phenomenon. In most situations, the wavelength is very short, however, so that diffraction and other wave-like behavior appear only with micro-domain-sized slits, as in the nickel crystal. Thus Thomson's experiments, with macro-domain size slits, did not reveal wave effects. The modern theory includes the possibility of forming micro-domain wave packets from electron waves. These tiny wave packets, which carry mass and electric charge, behave like particles in macro-domain experiments.

Figure 8.6 (to right). J. J. Thomson's "plum pudding" model for the atom. The tiny electrons (the raisins) repel one another but are held in place (in mechanical equilibrium) by a much larger "pudding" of positive charge. The positive charge has most of the mass of the atom but, unlike the electrons, is like a diffuse "cloud" spread throughout the volume of the atom. The electrons, if stimulated, can vibrate in place to produce light waves.



The MIP model. At the beginning of the twentieth century it was definitely established by experiment (electrolysis, the ability of matter to emit electromagnetic waves, and cathode ray studies) that matter could be divided into electrically charged components. Presumably, therefore, each electrically neutral atom had positive and negative parts. At last it was possible to identify how the particles interacted in the MIP model: the interactions were electric, transmitted by means of an electric field. Macro-domain chemical, phase and elastic energy were, therefore, really just electric field energy of the atoms and molecules. The mutual attraction of the positive and negative parts in adjacent atoms could explain the structure of molecules, crystals, and liquids. The relative motion of the electrically charged parts could explain the emission of electromagnetic waves (Section 7.3), including the spectra of visible light.

The negative constituents of atoms were the very light electrons, which Thomson had measured as less than one thousandth the mass of the atom as a whole. What was the positive part, which appeared to have almost all the mass of the atom?

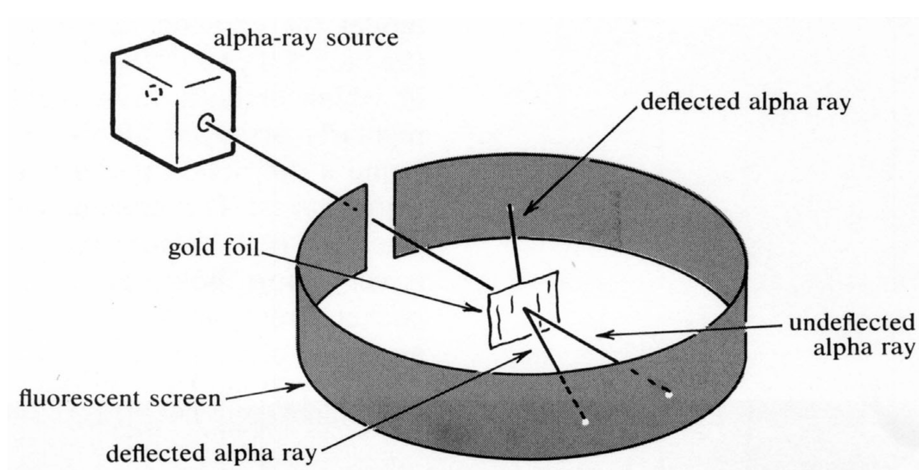
8.2 Early models for atoms

Thomson's model. J. J. Thomson himself was one of the first to propose an atomic model that took into account the atom's electrical nature (Figure 8.6). He suggested that an atom consists of a positive electric charge that is uniformly distributed within a spherical region of about

Figure 8.7 (below). Ernest Rutherford's scattering experiment. The alpha particles, which were known to have much more mass than an electron, emerge from the source in a narrow beam at high speed and strike the gold foil target. Most of the alphas continue undeflected through the foil in a straight line, but some are deflected at various angles. The physicist sits in darkness for long periods, counting the bright flashes on the circular screen created by the impacts of the deflected alpha particles.

Rutherford found that the number of alphas he counted at various angles did not match the predictions of the Thomson model (above). In particular, Rutherford was astounded (see quote on next page) to find a very few flashes showing that the alphas were occasionally bouncing <u>backwards</u>. He initially thought that such flashes could not be genuine.

It is extraordinarily unlikely that Thomson's plum pudding model (above) would bounce a heavy alpha particle backwards. Thus Rutherford invented another model, with the positive charge concentrated in a tiny but heavy <u>nucleus.</u>



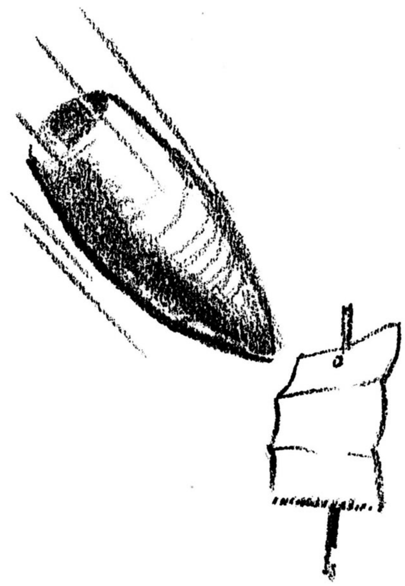


Figure 8.8 Analogue model for the collision of an alpha ray with an electron: a 15-inch shell hits tissue paper, as suggested by Ernest Rutherford in quote below to left. Would you expect the shell (or the alpha ray) to bounce back? Estimate the ratio of masses - this gives a rough estimate of the probability that the shell would bounce back.

Ernest Rutherford (1871-1937) was a New Zealander who came to England on a scholarship in 1895. After three years as a student of J. J. Thomson at Cambridge, he accepted a professorship at McGill University, Canada. Rutherford showed such virtuosity as an experimentalist that he was brought to Manchester as director of a research laboratory in 1907. At Manchester, Rutherford developed his nuclear model of the atom. An extraordinarily warm, perceptive, and stimulating man, Rutherford fostered and influenced an amazingly gifted group of young men, including Niels Bohr.

"It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you had fired a fifteen-inch shell at a piece of tissue paper and it came back and hit you." [see Figure 8.8 above]
Ernest Rutherford

the same size as an atom. The small, relatively light electrons were sprinkled throughout this region, somewhat like seeds in a watermelon or raisins in a plum pudding (Fig. 8.6). Since the negatively charged electrons repel one another, they tend to spread apart. The attraction of the positive charge, however, kept them from separating completely.

Thomson concluded that the electrons in this model should arrange themselves in ring-like layers. He also thought that atoms were held together in molecules and crystals by electrical forces between the unbalanced charges that are created when electrons are lost by one atom (which is then electrically positive) and are acquired by another atom (which is then electrically negative). In addition, Thomson tried, only partially successfully, to estimate how the electrons might vibrate so he could predict the frequencies of the emitted electromagnetic radiation.

Rutherford's experiment. An investigation conducted in Ernest Rutherford's laboratory in 1908 yielded a surprising result that could not be interpreted with the Thomson model for the atom. An extremely thin gold foil was bombarded with alpha rays, which are the disintegration products of the radioactive element radium (Fig. 8.7). From their interaction with magnets and electric charges, it was known that alpha rays are positively charged and are about 8000 times as massive as electrons. Most of the alpha rays that impinged on the foil went through without a measurable deflection, but a very small number were deflected by more than 90°, that is, back towards the source. Apparently, the deflected alpha rays hit something that was present here and there in the gold foil

In the Thomson model, the electrons were present "here and there." Could they have deflected the alpha rays? The very low mass of the electrons compared to that of the alpha rays meant that *none* of the alphas should be deflected by more than 90°, or backwards. *All* deflections from electrons should be small and none large (Fig. 8.8 and Figure 8.9a). But Rutherford's many experiments clearly showed that this was not true; therefore, he had to devise his own model for the way the electrons and the positive charge were arranged in the atom.

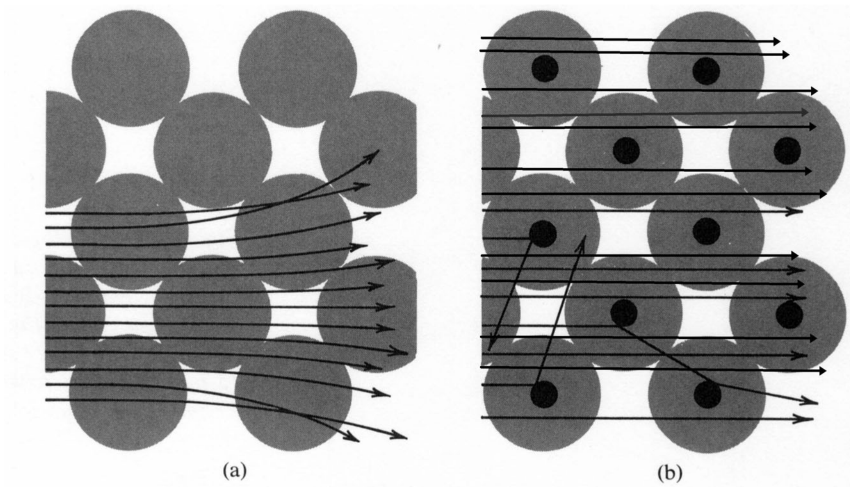


Figure 8.9 Deflection of alpha rays by a gold foil.
(a) As predicted by Thomson's "plum pudding model," most alphas would be deflected by small angles, but none would be deflected by more than 90°.
(b) As predicted by Rutherford's nuclear model, most alphas would be undeflected, but a few would actually strike the nucleus and thus be deflected by substantial angles. In addition, among alphas actually striking the nucleus, there would be a very few that would have "head-on" collisions. These alphas represent the very small, but not zero, number deflected by more than 90°

Rutherford's nuclear model. If the electrons weren't heavy enough to bounce the alpha rays back, what else in the atom could do it? Rutherford therefore thought further about the positive charge in the atom, which Thomson's model represented as being spread out like a cloud, but which included almost all the mass of the atom. Perhaps it was not spread out; perhaps it was actually concentrated with all its mass in a very small region of space. If alpha rays strike such a small, massive object head on, they would recoil in the direction from which they came or be deflected by large angles, as shown in Figure 8.9b. The alpha rays are like random "bullets" striking a target; only the ones that happen to hit the "bull's-eye" (the positive charge) would be deflected by large angles. Most of the alpha rays would simply continue in a straight line. The percentage of deflected alphas should be directly related to the size of the bull's-eye.

On the basis of this relationship Rutherford estimated the size of "bull's-eye," that is, the region occupied by the positive charge within the atom. Rutherford found that this region was much, much smaller than the size of a gold atom. Rutherford called this small, massive, positive charge the *nucleus* of the atom.

The atomic model invented by Rutherford, called the *nuclear atom,* has stood the test of time and, with many refinements, is still accepted today. However, when Rutherford proposed the nuclear model, there were many serious difficulties and various "mysteries" associated with this new model. We will now explain these problems and how they were resolved, which laid the groundwork for quantum theory.

Difficulties of the nuclear model. One of the mysteries was to explain how all the positive electric charge could remain clustered in a small region of space in spite of the enormous mutual repulsion of all the parts of the charge for one another. To solve this problem, it is necessary to make a model for the atomic nucleus itself, showing the parts it is composed of and how they interact (Section 8.6).

The problem of atomic stability. A second mystery was why the negative electrons, which were strongly attracted by the positive nucleus, remained outside the nucleus and were not pulled into it. The first attempt to resolve this mystery was not successful. Rutherford suggested that the atom is like a miniature solar system, with the light, negative electrons orbiting around the heavy, positive nucleus: the planetary model for the atom. Even though such an atom would not collapse immediately, the relative motion of negative and positive charges was known to lead to the emission of electromagnetic waves, as in Hertz's spark gap experiments (Section 7.3). These waves should be observable as light emitted by atoms and would gradually rob the atoms of energy until the electrons eventually collapsed into the nucleus (Fig. 8.10). Since atoms are not observed to emit light until they collapse (in fact, they are stable and do not collapse), the planetary version of the nuclear atom did not seem satisfactory.

The problem of line spectra. A third mystery was the line spectrum of light emitted by hot gases (Section 7.2). Hot gases emit light of selected frequencies, and the atomic models should give an explanation of the spectrum. The planetary nuclear atom led to the emission of light, as we have just explained, but the emission never stopped, and the frequency varied as the electron spiraled with an increasing frequency on its way toward the nucleus.

In spite of its difficulties, the nuclear atom also enjoyed successes. By studying the deflected alpha rays, Rutherford was able to determine the relative amounts of positive electric charge on various nuclei. He found that the positive charges on nuclei of different elements, such as carbon, aluminum, copper, and gold, varied in the same ratio as the atomic number of the element (Section 4.5). Since the negatively charged electrons in an atom must neutralize the positive charge on the nucleus, the number of electrons is just equal to the atomic number and varied from one element to the next in the periodic table. This startling result explained how the atoms of various elements differed from one another; furthermore, it explained the significance of the chemists' atomic number in terms of the structure of atoms.

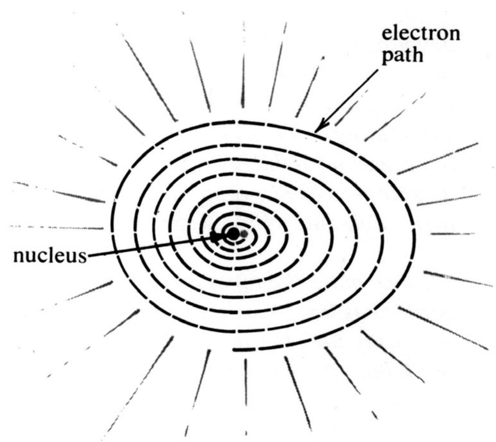


Figure 8.10 In the planetary model, the electron emits light and spirals into the nucleus as it loses more and more energy.

8.3 Bohr's model for the atom

Bohr's quantum rules:
1. Only certain selected ones of all conceivable orbits of the planetary motion of the electrons around the nucleus are permissible.
2. When an atom emits (or absorbs) light, it makes a quantum jump from one to another of the selected orbits. The energy transferred, called a quantum of energy, determines the frequency of the light, according to Eq. 8.1.

Equation 8.1

energy transferred = ΔE
frequency of light = f
Planck's constant = h

$$\Delta E = hf$$

Planck's constant h has a numerical value in the micro domain,

$$h = 6.6 \times 10^{-34}\ \text{kg-m}^2/\text{sec}$$

It therefore does not influence macro-domain phenomena directly.

" ... an electron of great velocity in passing through an atom and colliding with the [bound] electrons will lose energy in distinct finite quanta ... very different from what we might expect if the result of the collisions was governed by the usual mechanical laws."

Niels Bohr
Philosophical Magazine, 1913

Bohr's theory. Niels Bohr, a young Danish theoretical physicist temporarily working at Rutherford's laboratory, initiated a completely new approach to atomic models and achieved a decisive breakthrough in 1912. He accepted the findings of Thomson and Rutherford, but he did not submit their models merely to the laws of motion and electromagnetism as formulated by Newton and Maxwell, respectively. In addition, he was stimulated by the inconsistencies of the existing models to introduce new principles called *quantum rules* (stated to the left), and he found that these solved all the difficulties associated with the electron motion and light emission in the nuclear atom. Equation 8.1 is one of the quantum rules, called *Bohr's frequency condition.* The quantum rules and the procedures for using them make up Bohr's theory. When Bohr applied the quantum rules to the nuclear model for the hydrogen atom, which is the simplest of all atoms in that it has only one electron, he obtained excellent quantitative agreement with observations of the hydrogen spectrum. This explanation of the frequencies of the light emitted and absorbed by the hydrogen atom was a startling success of Bohr's theory.

A little later we will explain the details of the quantum rules and how to apply them. Now we will describe how they eliminate the critical difficulties of the nuclear model.

States of an atom. In any model for atoms, the state of an atom is described in terms of the arrangement and motion of its parts. Before Bohr, the radius of an electron's orbit around the nucleus was believed to become smaller and smaller as the electron lost energy by radiation and spiraled into the nucleus. The energy of the atom in any state was calculated from the speeds of the electron (kinetic energy) and the electric interaction among the electrons and the nucleus (electric field energy). With Bohr's rules, certain electron orbits and therefore certain states of atoms were selected, and all others were eliminated. The energy of each allowed state, however, was calculated from the kinetic and electric field energies as before. Each allowed state of an atom is called an *energy level* (Fig. 8.11).

Two important consequences follow from the existence of energy levels. First, there is an atomic state with the lowest permitted energy, called the *ground state.* An atom in the ground state cannot emit light, because emission of light would reduce its energy below the lowest permitted value. The ground state is, therefore, safe against collapse of the electrons into the nucleus. Second, when the atom is not in the ground state, it can emit light as the electron "jumps" from an orbit of higher energy to another of lower energy. The energy transferred to the light is a specific amount equal to the energy difference between the two states of the atom (Fig. 8.12). The frequency of the light can be calculated from the energy transfer according to the formula in Bohr's second rule and his frequency condition (Eq. 8.1). Hence the emission spectrum will consist only of the spectral lines arising from all the allowed orbital "jumps" of the electrons; the spectrum will not be a complete rainbow including all frequencies.

Niels Bohr (1885-1962) had just received his doctorate in Denmark when he went to England in 1911. He joined Rutherford's research group in Manchester, the "home" of the nuclear model of the atom. Bohr was intimately familiar with the latest physics theory (mostly developed in continental Europe, especially Germany, England's archrival) which revealed the stubborn contradictions and difficulties inherent in the nuclear model, and he had a remarkable capacity to connect with others as well as boundless enthusiasm for discussing the details of apparently conflicting ideas. The 26 year-old Danish theorist responded to the stimulation of Rutherford's lab with an extraordinary burst of creativity: audaciously inventing his own new "quantum rules" and using them, together with the latest theory, to resolve the well-known problems delaying acceptance of the nuclear atom. For this work, published in 1913 and 1915 (while Europe was entering World War I), Bohr was awarded the Nobel Prize in 1922.

Bohr's scientific genius was equaled by his courage, compassion, and wisdom. During World War II he helped rescue thousands of Jewish refugees by smuggling them out of Nazi-occupied Denmark to the safety of Sweden. Bohr's last years as Director of the Copenhagen Institute for Theoretical Physics were distinguished by unceasing efforts to secure international peace in an age threatened by nuclear holocaust.

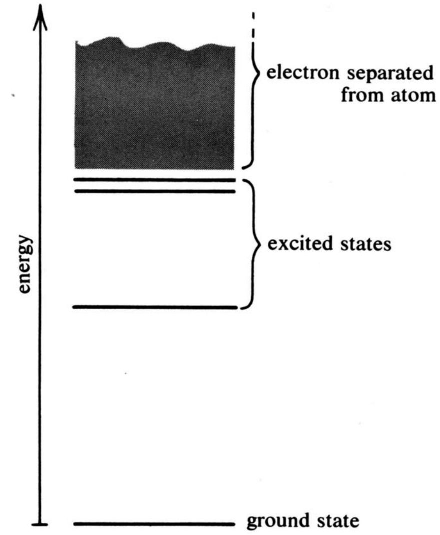


Figure 8.11 Diagram to represent the allowed states of an atom. Each line represents one state allowed by Bohr's quantum rules. The distance between lines represents the energy difference between the states. The diagram is called an energy level diagram.

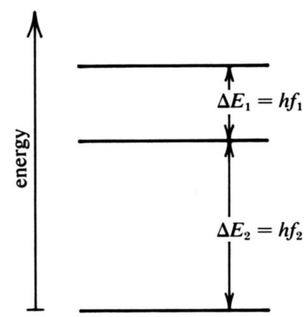


Figure 8.12 The frequencies of emitted spectral lines are found from the energy differences on an energy level diagram.

Revolutionary aspect of Bohr's rules. Bohr's theory provided a breakthrough because it added to the accepted laws and principles of physics. It did not merely combine the accepted laws to obtain a new conclusion. This is characteristic of a scientific revolution.

Bohr's rules do seem arbitrary and ad hoc, and they appear to conflict with other, well-established laws of physics. The question naturally arises why should Bohr's rules hold? In 1912 many physicists asked this question. However, a "why" question of this type is more a philosophical than a scientific question. The quantum rules are an integral part of Bohr's model: the rules must be accepted if the model is successful and rejected, or modified, if it fails. If a second successful model is available, we could choose the one that better satisfies our curiosity or preferences. But, in 1912, calculations based on Bohr's theory, and no other, produced sensible results that agreed precisely with the observed frequencies of light emitted and absorbed by the hydrogen atom. With only one model, we do not have a choice. Nevertheless, many of Bohr's contemporaries questioned the new theory, and several years passed before the quantum rules were accepted for their value.

Quantum ideas before Bohr. The idea of the quantum (and the associated idea that micro-domain systems had discrete, not continuous, energy levels, as shown in Figure 8.12) did not originate with Bohr. Max Planck (1858-1947) had introduced it into the theory of light emission from glowing solids (Section 7.2). Planck used the micro-domain model that represented a solid as a system of electrically charged oscillators (Section 7-3). To explain the data, Planck assumed that only selected states of the oscillators could radiate their energy and that the energies of these states were directly proportional to the oscillator frequency (Eq. 8.2). The formula relating the energy to the frequency depended on a new number, which Planck called the *quantum of action* (symbol *h*). This number (now called Planck's constant) and formula were similar to the ones introduced later by Bohr.

Equation 8.2

energy of oscillator = E
frequency of oscillator = f
Planck's constant = h

$$E = hf$$

"While this constant was absolutely indispensible to the attainment of a correct expression ... it obstinately withstood all attempts at fitting it, in any suitable form, into the frame of a classical theory."

Max Planck
Nobel Prize address, 1920

A few years after Planck's work, Einstein applied the quantum concept to the photoelectric effect (Fig. 8.13). When light is absorbed by certain metals, electrons may be ejected from the metal surface. The frequency of the light and the kinetic energy of the electrons were measured and compared. Below a certain frequency, no electrons were emitted. Above that frequency, the kinetic energy increased in a regular fashion when the light frequency was increased. Einstein found he could explain the experimental measurements in terms of energy transfer from the light to each ejected electron with the following assumption: radiant energy can be transferred to one electron only in certain definite amounts, called a quantum. The energy of one quantum of light is directly proportional to its frequency, with Planck's constant again making its appearance (Eq. 8.3).

Equation 8.3

energy of 1 quantum = E

$$E = hf$$

Quantum number—Applications of Bohr's quantum rules. Bohr used Planck's constant and the electron momentum for the application of his first quantum rule to specific cases. You will recall that the momentum of a particle is its inertial mass multiplied by its speed (Eq. 3.1). The allowed circular orbits (quantum states) of Bohr's first rule are selected with the requirement that the electron momentum multiplied by the circumference of the orbit is equal to a whole number times Planck's constant (Fig. 8.14 and Eq. 8.4). The whole number n that appears in Eq. 8.4 is called a quantum number. This is known as quantization in which some quantity (in this case, $2\pi r\mathcal{M}$) can have only certain restricted values that are multiples of a basic quantity (in this case h).

Equation 8.4

electron momentum = $\mathcal{M}$
orbit circumference = $2\pi r$
quantum number = n

$$2\pi r \mathcal{M} = nh$$
$$(n = 1, 2, 3, 4 \ldots)$$

Equation 8.4 restricts the radius and momentum of the electron to only certain allowed values. The radius r and the momentum $\mathcal{M}$ must have values that fit into one of the following:

$2\pi r\mathcal{M} = h$, or
$2\pi r\mathcal{M} = 2h$, or
$2\pi r\mathcal{M} = 3h$, and so on….

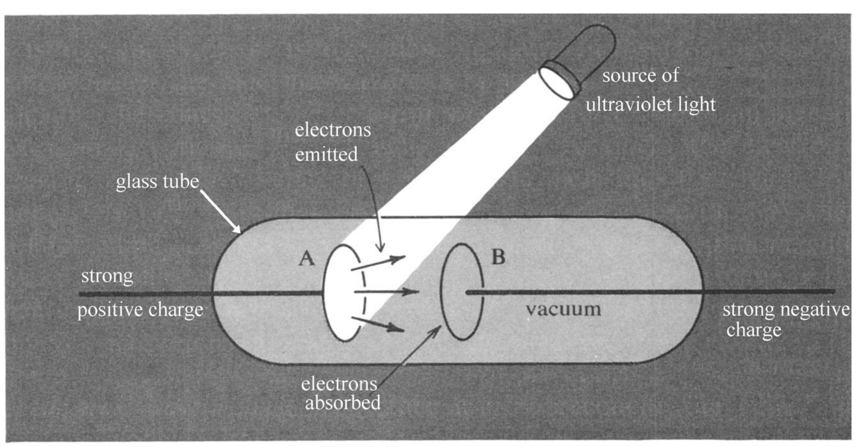


Figure 8.13 (to the right) The photoelectric effect. When ultraviolet light illuminates copper plate A, electrons are emitted by the copper plate, leaving it with a positive electric charge. The electrons are absorbed by copper plate B, giving it a negative electric charge.

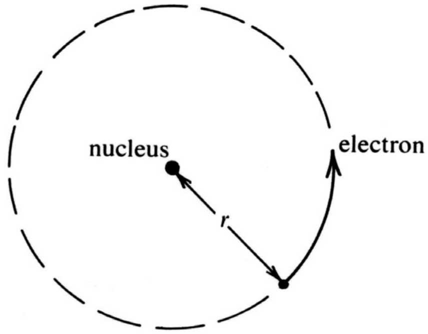


Figure 8.14 Electron in circular orbit around the nucleus. The radius of the orbit is r, the circumference is $2\pi r$.

In Bohr's second rule, the energy transferred when an electron jumps from one orbit to another is directly proportional to the light frequency (in (Eq. 8.1) in the same relation with Planck's constant as found by Planck and Einstein (Eqs. 8.2 and 8.3). The quantity of energy transferred can be expressed by means of the quantum numbers of the two orbits. As we stated above, this model predicts quite accurately the observed spectrum of the hydrogen atom. This outstanding success of Bohr's theory made it very attractive to many physicists in spite of its revolutionary nature, and intensive research to test its implications was undertaken.

Limitations of Bohr's theory. With the passage of time and accumulation of new, more precise data, Bohr's theory had to be modified in minor ways. However, more serious inadequacies were revealed in its applications to complex atoms containing many electrons. Bohr's theory led to the conclusion that all the electrons in the atom's ground state would occupy the same orbit close to the nucleus. Both the spectra and the sizes of atoms, however, indicated that this did not happen, but that electrons arranged themselves in different orbits, some close to the nucleus, some farther away. Furthermore, the repulsive interaction of the electrons with one another could not be included satisfactorily in the theory. The difficulty of treating systems containing many electrons made it impossible to apply the model to the formation of molecules out of atoms, because all molecules do include many electrons.

In addition, the model suffered from the philosophical weakness that it combined the old laws of Newton and Maxwell with the new quantum rules. This rather capricious combination was unattractive for some physicists, who could not see clearly just when the old and when the new was to be employed. Nevertheless, Bohr's theory was the first step, the breakthrough, of a scientific revolution in the theories and models for micro-domain physical systems. This revolution, which reached its climax in the nineteen twenties, was throughout inspired and guided by Bohr.

8.4 The wave mechanical atom

In Bohr's mysterious quantization, the electron momentum multiplied by the orbit's circumference is equal to a whole number times Planck's constant (Eq. 8.4). Another way to state this condition is that the length of the path of the electron in its orbit ($2\pi r$) is a whole number times Planck's constant divided by the electron momentum (Eq. 8.5)

Electron waves. Now, this relation reminded the French physicist Louis de Broglie of the condition for the wavelength of waves permitted in a tuned system (Eq. 8.6; see Eq. 6.4). Using Bohr's model for the atom, de Broglie interpreted the circumference of the orbit as the length of the tuned system. By comparing Eqs. 8.5 and 8.6 and identifying the circumference, $2\pi r$, with L, we can see the close similarity between $h/\mathcal{M}$ and the wavelength, λ. De Broglie therefore introduced the idea that the motion of electrons is governed by a wave whose wavelength is inversely proportional to the electron momentum (Eq. 8.7a) and whose wave number ($k = 1/\lambda$) is therefore directly proportional to the momentum (Eq. 8.7b). An atom is thus viewed as a tuned system that contains standing waves of one or more electrons. The electrons are refracted around the positively charged nucleus by its electrical field, which affects electron waves in the same way that a change of index of refraction affects light waves (Fig. 8.5).

Equation 8.5

$$2\pi r = n\,(h/\mathcal{M})$$

$(n = 1, 2, 3, ...)$

Equation 8.6

length of tuned system $= L$
wavelength $= \lambda$
number of waves $= n$

$$L = n\lambda$$

Equation 8.7

wave number $= k$

$$\lambda = h/\mathcal{M} \quad \text{(a)}$$

$$k = \mathcal{M}/h \quad \text{(b)}$$

Free electrons. If the motion of electrons is really governed by waves while they orbit the nucleus, the same should be true for electrons in a cathode-ray beam. Such electrons should therefore be diffracted by a suitable "grating" whose slits are separated by a distance comparable to the wavelength as given in Eq. 6.13. A nickel crystal provides such a grating. And, as we pointed out in Section 8.1, Davisson and Germer accurately verified this prediction of de Broglie's model.

Wave mechanics. The branch of physics in which the motion of electrons is represented as wave propagation is called wave mechanics. Electron waves are diffracted by obstacles. They are refracted when passing through the electric fields of the nucleus and of other electrons. The computational method for finding electron standing wave patterns was developed by the Austrian physicist Erwin Schrödinger and has been successfully applied to simple atoms and molecules. In principle, wave mechanical calculations make it possible to compute the physical and chemical properties of all substances. In practice, the calculations could be done for only the simplest atoms. At that time, the calculations for atoms with more than 2 or 3 electrons were much too complicated; now high-speed computers have successfully calculated the properties

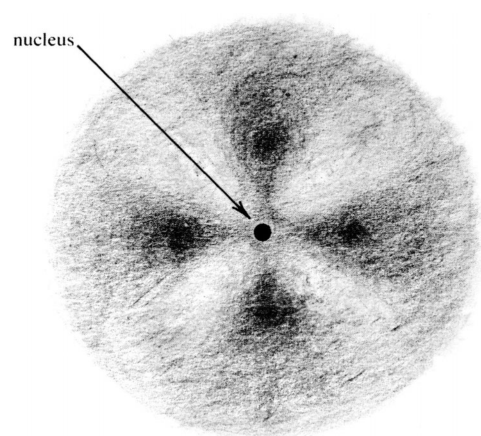


Figure 8.15 Diagram showing electron waves refracted around the nucleus. Four wave maxima and four minima are shown. Compare with Fig. 8.14, which represents the orbit of electron particles.

of many atoms and molecules; thus confirming the validity of the wave mechanical model to extremely high accuracy.

Electron waves in atoms and molecules. According to the wave mechanical model, therefore, an atom consists of a small nucleus surrounded by a standing wave of one or more electrons. The atom is very much larger than the nucleus, with most of its space occupied by the electron waves. The size of the atom is approximately equal to one or a few wavelengths of the electron waves forming a standing wave pattern. The atomic collapse predicted in the particle model for the electron does not occur because the particle description is not applicable. The separation of adjacent atoms in a molecule or crystal is maintained by the mutual repulsion of the negatively charged electron waves surrounding the nuclei of adjacent atoms. The formation of molecules is explained by the partial merging of certain electron waves, called valence electrons, for which the attraction by the adjacent nucleus is stronger than the repulsion by the adjacent electron wave.

Interaction of atoms with light. The wave mechanical model retains the concept of quantum jumps in which the electron gains or loses energy when the atom absorbs or emits light. In a quantum jump, the electron standing wave patterns are transformed from those of the initial state to those of the final state. The frequency of the light, still determined by Bohr's frequency condition (Eq. 8.1), is directly proportional to the energy gain or loss. Schrödinger's mathematical formulation, however, makes it possible not only to calculate the frequency but also to relate the intensity of the emitted light to the standing wave patterns in the atom's initial and final states.

Louis Victor de Broglie (1892-1987) is unique among great physicists because he became interested in physics at an age when most physicists have completed their major work, and even this late start was interrupted by service in the army during World War I. After the war, in his brother's scientific laboratory, the idea that electrons may be considered to have wave properties occurred to him. This theory, delivered in his doctoral thesis, Recherches sur la Theorie des Quanta, in 1924, won de Broglie the Nobel Prize in 1929.

Erwin Schrödinger (1887-1960) was born and educated in Vienna. Violently anti-Nazi, Schrödinger barely escaped a concentration camp by fleeing Austria in 1940. Most of his last years were spent as director of the Institute for Advanced Studies in Dublin, Ireland, where his interest was attracted to questions related to the nature of life. In 1924 and 1925 Schrödinger developed wave mechanics using de Broglie's idea that stationary states in atoms correspond to standing matter waves. For this distinguished work, Schrödinger shared the Nobel Prize with P. A. M. Dirac in 1933.

8.5 Wave particle duality and the uncertainty principle

You may wonder how it is possible to reconcile the particle and wave models for the electron. Cathode rays were originally invented as the intermediaries in an interaction-at-a-distance between the cathode and the glass of the discharge tube. When it was found that they possessed electric charge and inertial mass, a particle model became accepted; the particles were called electrons. Later, the diffraction experiments gave evidence of the electron's wave nature.

The answer to the question "Where is the electric charge of the electron located?" illustrates the apparent contradictions between the two models. In the particle model, the charge is located at the position of the particle. In the wave model, the charge is spread throughout the region occupied by the wave; the electron is found more often (and the charge is more concentrated) where the wave has a large amplitude and the electron is found less often (and the charge is less concentrated) where the wave has a small amplitude. It should be possible, you may say, to find the correct model by means of an experiment. Unfortunately, as we will explain below, the experiments that can be carried out do not help to distinguish between the two models. As a matter of fact, there was a famous years-long debate between Einstein and Bohr, in

The letters below illustrate the long-running debate between Bohr and Einstein about the probability interpretation of quantum theory. In 1949 (when quantum theory, including the probability interpretation, had been well accepted by essentially the entire physics community) Bohr sent a letter to Einstein wishing him a happy 70th birthday.

Einstein wrote back, thanking Bohr for his good wishes and added, "This is one of the occasions, which is not dependent on the disquieting question if God throws the dice or if we should hold on to the available physical description of realities."

Bohr then wrote back, ". . .To continue in the same jocular tone, I have no choice but to say on this painful issue that it really doesn't matter in my opinion whether or not we should hold on to the physical description of accessible realities or further pursue the path you have shown and recognize the logical assumptions for the description of realities. In my impertinent way I would suggest that nobody—not even God himself—would know what a phrase like playing dice would mean in that context."

"Like following life through creatures you dissect, You lose it in the moment you detect."

Alexander Pope
Essay on Man, 1732

which Einstein tried to find weaknesses in the quantum theory, but in which (in the opinion of most physicists today) Bohr actually found the weaknesses of Einstein's objections.

The effect of measurement on the state of a system. Consider the following simple example. A blind person tries to find the location of a glass marble on a table. He feels with his hands, measuring from the edges of the table, until his fingers touch the marble. Then he knows where he found it, but he will also have bumped it slightly and nudged it to a new location.

In any real experiment, some "nudging" of the system under study always takes place. What we tend to do, however, is to idealize by imagining a very, very gentle "nudge"—just enough, as it were, to tell the blind person that he has found the marble, but not enough to displace it. You can readily see that this ideal does not really exist, that finding the marble will always influence the marble, and observing a system will always influence or disturb the system in some way.

Even though the scientist has much more delicate instruments than the blind man's hands, the quantization of energy limits the "gentleness" with which he can operate. When he detects the position of the electron, there has to be some interaction and energy transfer between the electron and the measuring instrument. The smallest amount of energy that can be transferred is 1 quantum. Since the electron's inertial mass is very small, even the transfer of a very small amount of energy greatly affects its state.

Probability interpretation. The structure of the quantum theory, in other words, makes it impossible to carry out operations that identify an electron definitely as a wave or a particle. When you treat the electron as a particle and direct it through a small opening, the resulting diffraction pattern reminds you of its wave nature. When you treat the electron as a wave and try to separate a portion of the wave pattern from the remainder, you find that you get either all of it or none of it, as you would for a particle. The question "Is the electron a wave or a particle?" therefore does not have an operational meaning within the theory.

Nevertheless, there is a relationship between the two views. The amplitude of the electron wave can be related to the probability for locating the charge of the electron. Suppose, for example, an investigator makes many measurements of the position of the electron's charge in the ground state of an atom. He then finds the charge frequently (that is, with high probability) in regions where the wave has a large amplitude (Fig. 8.15). He finds the charge rarely (that is, with low probability) in regions where the wave has a small amplitude. He has to make many observations, however, and each time he finds all or none of the charge of the electron at the position he is observing.

Electron wave packets. Another way to reconcile the wave and particle models is to represent the electron as a wave packet. A wave

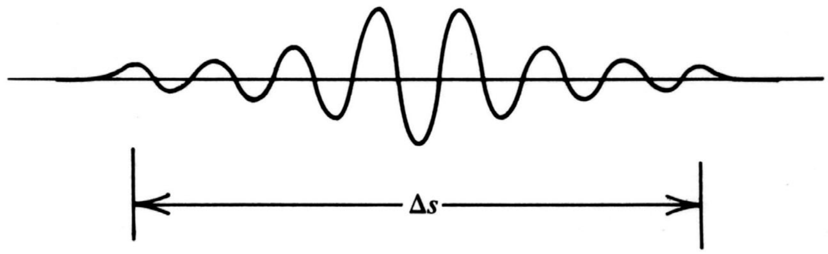


Figure 8 16 An electron wave packet of size As.

Equation 8.8

size of the wave packet $= \Delta s$
uncertainty in wave number $= \Delta k$

$$\Delta s\ \Delta k = 1$$

Equation 8.9

uncertainty in wave number $= \Delta k$
uncertainty in momentum $= \Delta \mathcal{M}$

$$\Delta k = \Delta \mathcal{M} / h$$

Equation 8.10

$$\Delta s\ \Delta \mathcal{M} = h$$

You must remember that a particle is an idealized object used in the construction of models. A particle occupies a point in space and has inertial mass, speed, and momentum. A wave packet is an alternative, somewhat more complex, idealized object which we use instead of a particle in constructing models in the micro domain.

packet differs from a single wave train in that it does not extend throughout space. It differs from a particle in that it is not localized at one point. Wave packets are described by an uncertainty principle (Eq. 8.8), which was explained in Section 6.2. This principle applies to electron waves also.

We will now interpret the uncertainty principle for electron wave packets. The size Δs of the wave packet (Fig. 8.16) is interpreted as the uncertainty in the position of the electron. The uncertainty Δk in the wave number is interpreted as an uncertainty in the momentum of the electron (Eq. 8.9, derived from de Broglie's relation Eq. 8.7b). You can therefore infer an uncertainty relation between the position and the momentum of the electron (Eq. 8.10). The smaller the uncertainty in the position, the larger the uncertainty in momentum. And vice versa. It is not possible to find the precise position and momentum of an electron simultaneously. This property of an electron means that it cannot be described by a particle model.

Matter waves. Experiments with beams of alpha rays, atomic nuclei, and even entire atoms and molecules have shown that these, too, are diffracted; in other words, their motion is in accord with the theory of wave propagation. You may now ask, "Where does this end? Does all matter exhibit wave properties?" The answer is, in principle, yes. A wave packet is associated with every material system. In practice, however, the wavelengths of systems in the macro domain are so small that diffraction effects are undetectable for them. The uncertainty principle, for example, reveals that a speck of dust with a mass of 1 milligram can be treated as a "particle" with negligible uncertainty of position and momentum, even though an electron cannot be treated as such (Example 8.1). The reason is the micro-domain numerical value of Planck's constant.

EXAMPLE 8·1 Applications of the uncertainty principle using Eq. 8·10, $\Delta s \approx h/\Delta \mathcal{M}$.

(a) Electron:
mass $M_I = 10^{-30}$ kg
speed $v = 10^{7}$ m/sec

momentum $\mathcal{M} = M_I\, v = 10^{-30}\ \text{kg} \times 10^{7}\ \text{m/sec} = 10^{-23}\ \text{kg-m/sec}$

ten percent uncertainty in momentum:

$$\Delta \mathscr{M} = 10^{-24} \text{ kg-m/sec}$$

uncertainty in position:

$$\Delta s \approx \frac{h}{\Delta \mathscr{M}} = \frac{6.6 \times 10^{-34} \text{ kg-m}^2\text{/sec}}{10^{-24} \text{ kg-m/sec}} = 6.6 \times 10^{-10} \text{ m}$$

(approximately five times the diameter of a hydrogen atom)

(b) Dust particle:
mass $M_I = 10^{-6}$ kg
speed $v \ = 1$ m/sec

$$\text{momentum } \mathscr{M} = M_I v = 10^{-6} \text{ kg-m/sec}$$

one-tenth percent uncertainty in momentum:

$$\Delta \mathscr{M} = 10^{-9} \text{ kg-m/sec}$$

uncertainty in position:

$$\Delta s = \frac{h}{\Delta \mathscr{M}} = \frac{6.6 \times 10^{-34} \text{ kg-m}^2\text{/sec}}{10^{-9} \text{ kg-m/sec}} = 6.6 \times 10^{-25} \text{ m}$$

(a completely negligible uncertainty)

8.6 The atomic nucleus

In the preceding section, the atomic nucleus was described as a massive, positively charged particle whose electric field refracted the electrons into a standing wave around the nucleus. As a matter of fact, scientists have formulated models for the nucleus itself as a complex system of interacting parts, a system capable of acting as energy source or energy receiver by changing its state.

Radioactivity. Already before Rutherford's invention of the nuclear atom, mysterious rays (including Rutherford's alpha rays) had been observed by Henri Becquerel to affect photographic plates near certain materials, especially compounds containing uranium, and had been given the name radioactivity. These rays were soon identified by their interaction with a magnetic field as having three components (Fig. 8.17): massive, positively charged alpha rays (Rutherford's tool for discovering the nucleus); light, negatively charged beta rays (later identified as electrons); and electrically uncharged gamma rays, which acted similarly to X rays (later identified as electromagnetic radiation). Scientists studied the rays and the sources intensively, finding that certain elements must be the source of the rays and that, amazingly enough, these elements seemed to be "decaying" into new elements (the objective of the alchemists' centuries-old, but unsuccessful, quest). Initially, because the rate of all known chemical reactions depended on temperature, it also seemed obvious that the rate at which the rays were emitted should depend on temperature, or on the specific other elements with which the radioactive elements were combined. However, many studies established conclusively the opposite: the rate of radioactivity depended only on the amount of the particular radioactive elements present, and neither on the temperature nor anything else.

Radioactive elements each have a characteristic half-life (the time it takes for half of a sample of the element to decay into another element, leaving one half of the original element). The half-life of radium is about 1600 years, and uranium's is almost five billion years! On the other hand, the half-life of other radioactive elements is a tiny fraction of a second.

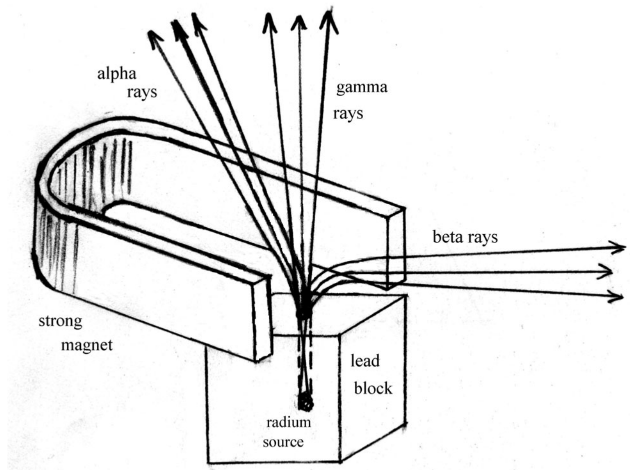


Figure 8.17 Becquerel's separation of alpha, beta, and gamma rays by using a magnetic field. Compare the deflections with those of cathode rays in Figure 8.4.

Henri Becquerel (1852-1908) studied at the Ecole Polytechnique, where he was appointed a demonstrator in 1875 and a professor in 1895. His accidental discovery of radioactivity in 1896 triggered an avalanche of research into radioactivity and its source in the atomic nucleus.

" I particularly insist on the following fact, which appears to me exceedingly important and not in accord with the phenomena which one might expect to observe: the same encrusted crystals placed with respect to the photographic plates in the same conditions and acting through the same screens, but protected from the excitation of incident rays and kept in the dark, still produce the same photographic effects."
Henri Becquerel
Comptes Rendus, 1896

The discovery of helium in radioactive ores and the experimental similarity of alpha rays to electrically charged helium atoms led Rutherford in 1903 to the conclusion that radioactivity was a breakdown of an atom of one element into an atom of another element. Uranium was gradually converted to radium and finally to lead, with alpha rays being emitted. The alpha rays then form helium atoms, which accumulate in the material.

This was the first serious attack on the concept that atoms were indivisible and immutable. (The next major step was Rutherford's invention of the nuclear atom.) The electric charges, masses, and energies observed in radioactivity quickly led Rutherford to the further conviction that radioactivity was a breakdown of the nucleus itself and did not involve the atomic electrons in a significant way. From the very beginning of the nuclear model it was clear that the nucleus was divisible and changeable. It was natural, therefore, for Rutherford to be puzzled that the positive charges in the nucleus held together as well as they did. Because of the mutual electrical repulsion of the positive charges, he wondered why every atomic nucleus did not disintegrate, as radioactive nuclei in fact do when they emit alpha rays.

Modern nuclear models. According to the currently accepted model, the atomic nucleus consists of positively charged constituents called *protons* and electrically neutral constituents called *neutrons,* all refracted into a standing wave pattern by their interaction. Neutrons and protons have approximately the same mass, but each is about 2000 times as massive as an electron. The nucleus of a hydrogen atom is one proton, an alpha ray consists of two protons and two neutrons, an oxygen nucleus includes eight protons and eight neutrons, and a uranium nucleus contains 92 protons and 143 to 146 neutrons.

Atomic number. The positive electric charge of the proton is

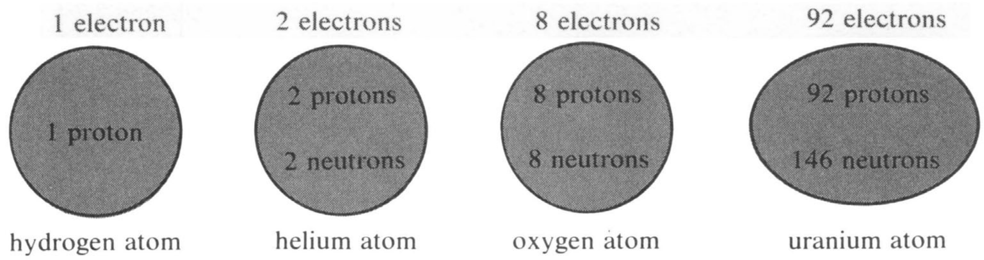


Figure 8.18 (above) The composition of four electrically neutral atoms. The numbers of electrons and protons are always equal.

equal in magnitude to the negative electric charge of the electron. Therefore, the hydrogen atom composed of one proton and one electron is electrically neutral. The number of protons determines the total positive electric charge of the nucleus, and this, in turn, determines the number of electrons that are bound in the neutral atom (Fig. 8.18). The number of electrons, finally, determines the chemical properties of an element (Section 8.2). The chemical properties, in this model, are therefore traceable to the number of protons in an atomic nucleus of the element. The number of neutrons in the nucleus does not play a role in an element's chemical properties but it does contribute to the mass of the nucleus and to its stability.

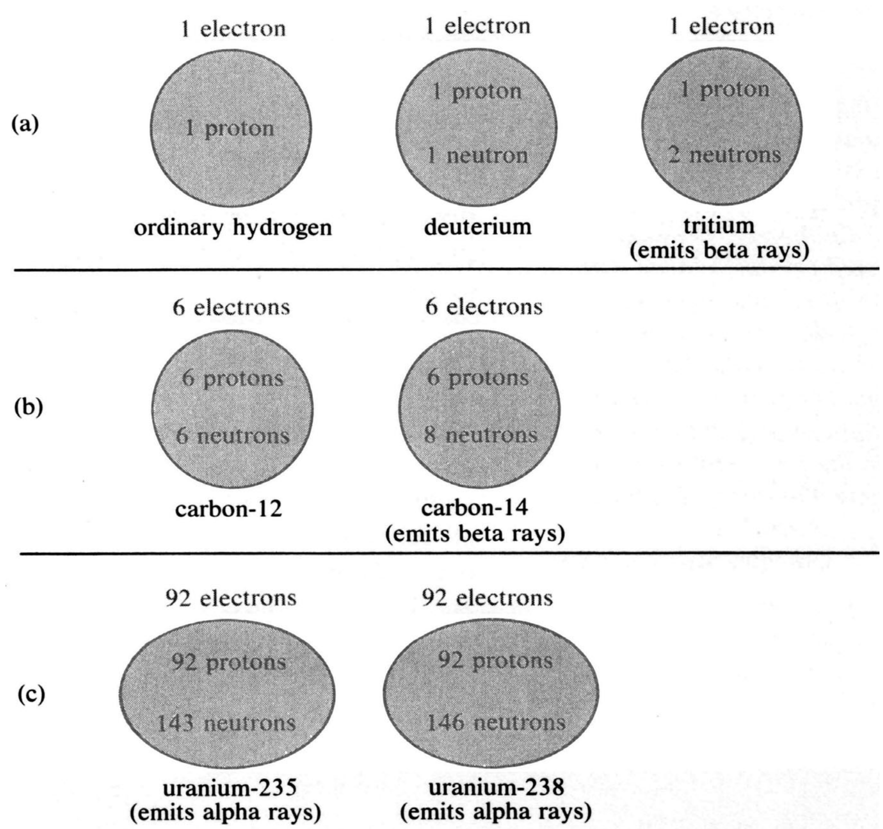
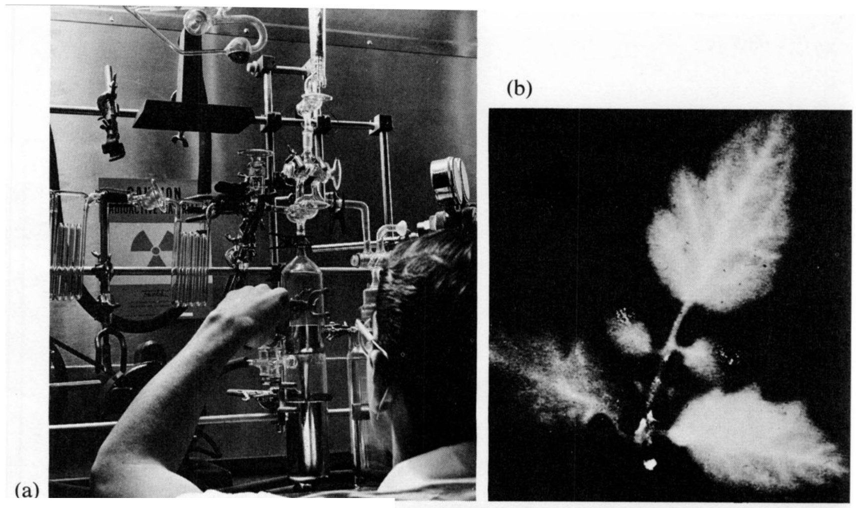


Figure 8.19 Diagrammatic representation of isotopes.
(a) Isotopes of hydrogen.
(b) Isotopes of carbon.
(c) Isotopes of uranium.



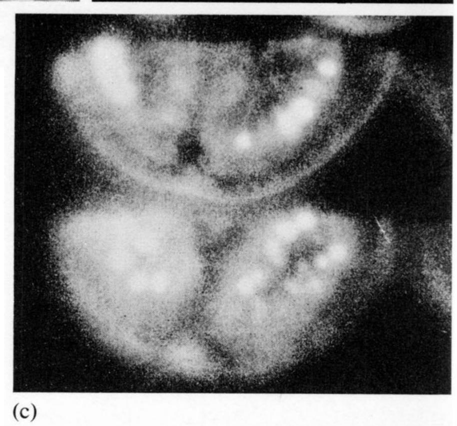


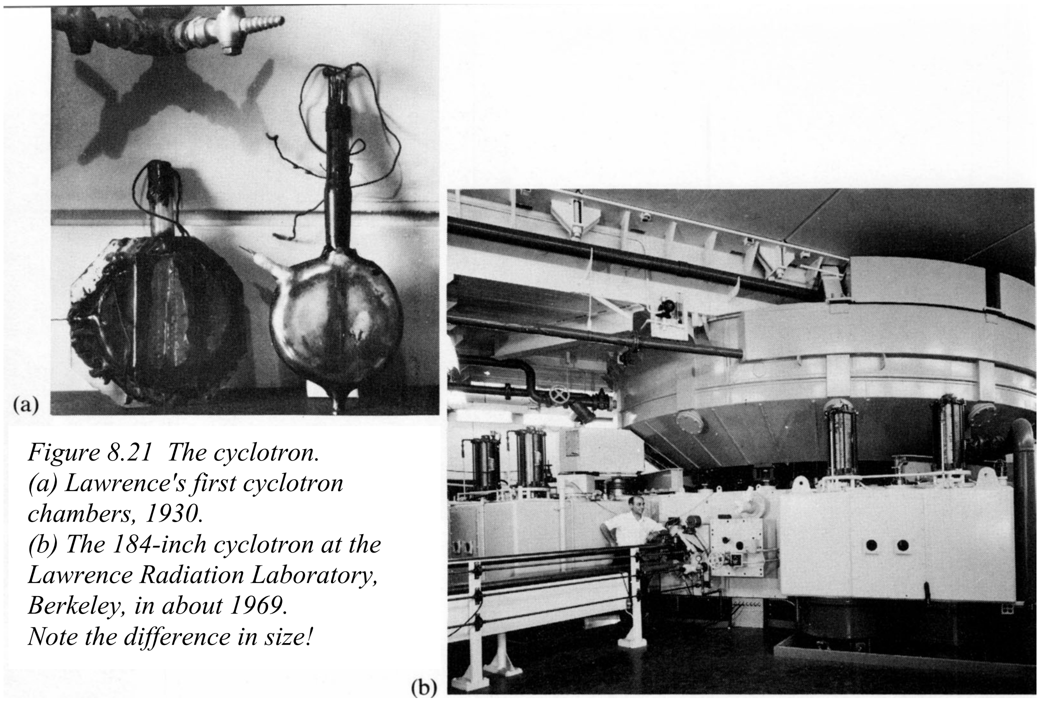
Figure 8.20 Application of radioactive isotopes.
(a) Radioactive carbon is being used in this drug metabolism study. Drugs labeled with radioactive materials yield information about concentration of compounds in tissue and about molecular structures, and provide clues to biosynthesis and interactions of normal body chemicals.
(b) The first "radio autograph" made in the dark with an artificial isotope of phosphorus. The isotope, produced in E. O. Lawrence's cyclotron, shows its presence in the leaf by affecting the photographic plate on which the leaf was placed.
(c) Radio autograph of a sliced tomato fed with a trace of radioactive zinc. The lightest spots show where the radioactivity is strongest. What can you infer about the role of zinc in growing tomatoes?

Isotopes. The name "isotope" is given to atoms of one element whose nuclei differ only in their number of neutrons (Fig. 8.19). Most elements have one or a few known stable isotopes and a few radioactive isotopes, which emit beta rays. The radioactive isotopes are thereby transformed into stable isotopes of another element. Isotopes have found many uses in industry, science, and medicine (Fig. 8.20).

In the high temperatures at the center of stars atomic nuclei become separated from their electrons. Thus, the nuclei are not surrounded by mutually repelling electron waves; therefore, the nuclei can interact with each other when they collide. In fact, such collisions result in the large nuclear energy release that, in turn, maintains the high temperature of the sun and stars.

Nuclear stability. Like electrons, both neutrons and protons exhibit wave-particle duality and obey the uncertainty principle. Most important, protons and neutrons participate in a non-electrical attractive interaction-at-a-distance, which binds them to form stable nuclei. The interaction, called the *nuclear interaction,* is much stronger than the electrical repulsion for inter-particle distances less than about 10^{-15} meters, but it becomes exceedingly weak for larger distances. The nuclear interaction acts over such extremely small distances that not even the nuclei of adjacent atoms in solids or liquids are affected. Consequently, there are no macro-domain manifestations of the nuclear interaction and it does not play a role in everyday phenomena.

The role of the neutrons in the nucleus appears to be one of stabilizing the nuclear system. They participate in the strong attractive



(a) (b)

Figure 8.21 The cyclotron. (a) Lawrence's first cyclotron chambers, 1930. (b) The 184-inch cyclotron at the Lawrence Radiation Laboratory, Berkeley, in about 1969. Note the difference in size!

nuclear interaction and not in the mutual electrical repulsion of the protons. Stable nuclei contain a number of neutrons about equal to or somewhat larger than the number of protons (Figs. 8.19 and 8.20). Nuclei that deviate from the ideal are radioactive and disintegrate by the emission of alpha or beta rays to form more stable nuclei.

Nuclear reactions. *The accelerator* is a research tool that has made possible the systematic study of atomic nuclei and their properties. An accelerator produces a beam of protons, electrons, or alpha rays with very high kinetic energy. The first of these machines was the *cyclotron* invented by E. O. Lawrence in 1931 (Fig. 8.21). Since then, new families of accelerators have been designed and built to study the properties of nuclei during highly energetic collisions.

Ernest O. Lawrence (1902-1958), the father of the modern accelerator, was professor of physics at the University of California at Berkeley and the first director of the university's famed Radiation Laboratory. For his researches in atomic structure, development of the cyclotron, and its use in artificially induced radioactivity, Lawrence won the Nobel Prize in 1939. During World War II, he was one of the chief participants in the race to develop the atomic bomb before the Germans

When the energetic rays produced by an accelerator interact with the nuclei in a target such as thin aluminum foil, they set off a series of changes in the state of the target nucleus. These changes are called *nuclear reactions* (analogous to chemical reactions). The result is the formation of new nuclei with a changed number of neutrons and protons. Frequently these nuclei are radioactive isotopes and are useful in scientific research and in medicine.

Energy transfer. One of the most important practical outcomes of the study of radioactivity and nuclear reactions was the discovery that nuclear transformations involve very a large energy transfer from nuclear field energy to kinetic energy of the reaction products. This is easy to understand in the light of the model we have presented above. The nucleus consists of very strongly interacting neutrons and protons. When the standing wave pattern of these is changed, the nuclear energy stored in the system is changed, with the energy difference being transferred

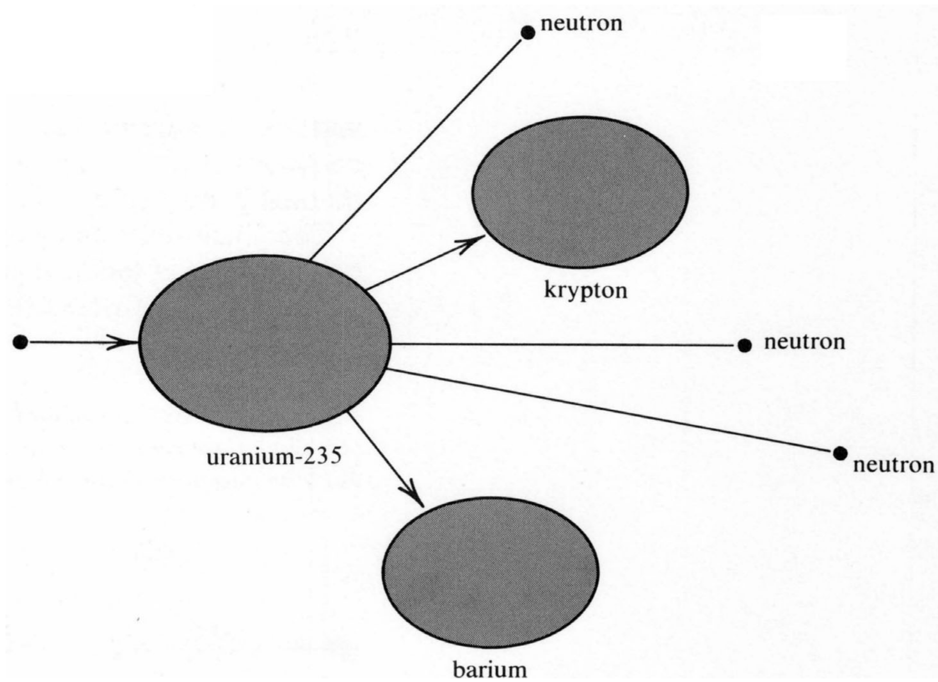


Figure 8.22 Diagram of the nuclear fission process. The isotope uranium-235 is especially susceptible to fission after being struck by a neutron.

to other forms, such as kinetic energy of the nuclear reaction products.

Interestingly enough, the change in energy of a nucleus (as determined by the energy transfer to other forms) also manifests itself as a change in the mass of the nucleus. The energy transfer and the mass change are directly proportional (Eq. 8-11), as predicted by Einstein in his special theory of relativity in 1905.

Equation 8.11
energy transfer = ΔE
change in mass = ΔM_I
speed of light = c

$$\Delta E = \Delta M_I c^2$$

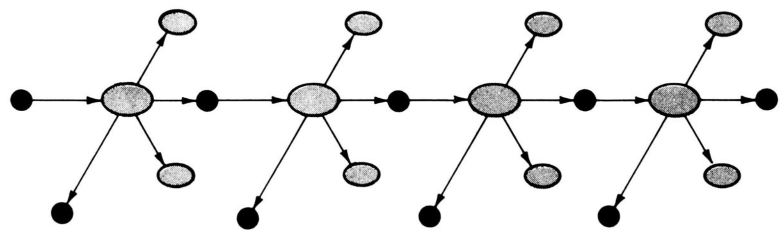
"If a body gives off the energy E in the form of radiation, its mass diminishes by E/c^2. The fact that the energy withdrawn from the body becomes energy of radiation evidently makes no difference, so that we are led to the more general conclusion that the mass of a body is a measure of its energy-content."

Albert Einstein
Annalen der Physik, 1905

Nuclear fission. *A* nuclear reaction in which an especially large amount of nuclear energy is released is the process of *fission.* Very large nuclei, such as uranium, are capable of fission when they are bombarded by neutrons. In the fission process, a neutron interacts with the uranium nucleus, which then disintegrates into two very energetic nuclear fragments (of about half the mass of uranium) and two or three neutrons (Fig. 8.22). The nuclear fragments transfer their kinetic energy to atoms with which they collide and thereby increase the temperature of the material in which the fissioning nucleus was embedded. In this process, nuclear energy is transformed into thermal energy, with the fission fragments acting as intermediate energy receivers and sources.

Nuclear chain reactions. The utilization of nuclear fission for macroscopic energy transfer requires the fissioning of large numbers of uranium nuclei. This has been accomplished successfully by using the principle of the *chain reaction.* One neutron is required to trigger the fission process, and two or three neutrons are produced. If one of the latter neutrons is allowed to fission a second uranium nucleus, and one of those produced then fissions a third, and so on, the process continues with the result that much nuclear energy is converted to thermal energy (Fig. 8.23a).

The chain reaction can lead to a nuclear explosion, as in an atomic bomb, if two of the neutrons produced during a fission process trigger another fission process (Fig. 8.23b). Then one fission is followed by



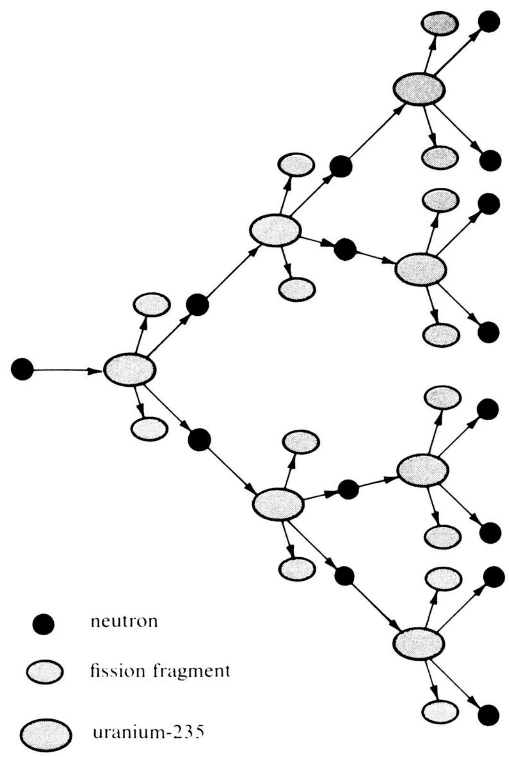


Figure 8.23 Self-sustaining chain reaction. The neutrons are represented by the small black dots. After a neutron strikes a uranium nucleus, the nucleus is likely to split (fission) into two smaller nuclei and emit other neutrons. These neutrons can be lost (leave the sample or be absorbed by impurities), or they can strike another uranium nucleus.

(a) Non-explosive chain reaction, as in a nuclear power reactor. If some neutrons are lost, so that exactly one neutron from each fission causes another fission, the number of chain reactions is maintained, and the amount of energy released stays constant.

(b) Explosive chain reaction, as in a bomb. If the loss of neutrons is minimized, so that more than one neutron from each fission causes another fission, the number of fission reactions multiplies, and the energy released grows rapidly without stopping. In this diagram, two neutrons from each fission are shown, each of which causes another fission reaction. Thus the number of reactions doubles in each "generation." To see how extraordinarily fast such growth can be, calculate how much money you would have after a month if you earned one penny on the first day, two pennies on the second day, and so on.

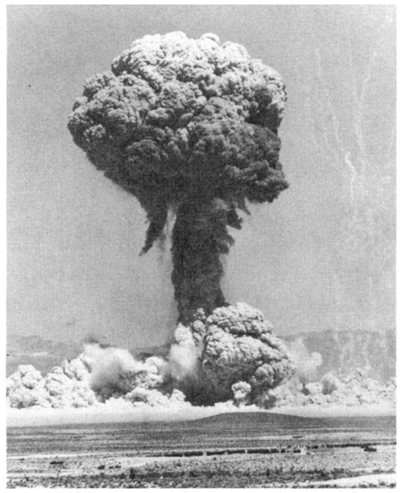


Figure 8.24 Nuclear explosion, Nevada, 1955.

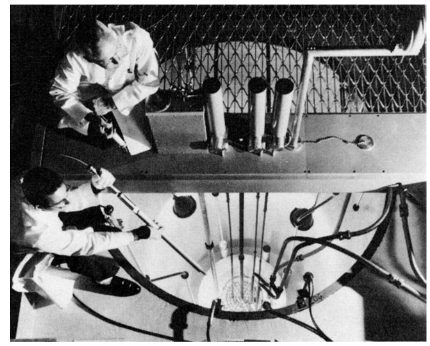


Figure 8.25 In this nuclear reactor for laboratory research, the uranium-containing reactor core is at the bottom of a 24-foot deep tank filled with water. Control rods and measuring devices extend down from the top.

two, the two by four, the four by eight, and so on, until (after about 70 steps) all remaining nuclei fission at once with an enormous energy release equivalent to thousands of tons of TNT (Fig. 8.24).

Nuclear reactors. To harness the chain reaction, it is necessary to have exactly one neutron propagate the process, not more and not less. An excess of neutrons leads to an explosion, a deficiency to extinction of the chain reaction. *A nuclear reactor* is a system in which a steady chain reaction is achieved by careful control of the neutron economy (Fig. 8.25). Some neutrons escape from the system, some neutrons are absorbed by structural members of the reactor, and some neutrons are absorbed without causing fission by cadmium or boron *control rods.* Everything is placed so that exactly one neutron per fission process sustains the reaction. As uranium in the reactor is depleted, the control rods are gradually withdrawn to compensate for the remaining uranium's lowered efficiency. When the reactor is to be shut down, the control rods are inserted more deeply. A control mechanism that carefully monitors the level of neutron production supplies negative feedback via the control rods to counteract any deviations from the desired operating level.

Summary

The many-interacting-particle model for matter had been proposed, discarded, and resurrected several times since the days of the Greek philosophers, when Dalton directed attention at the ratios of weights and volumes in which elements combine to form compounds. Dalton's model was so useful in correlating chemical data that it was accepted after a few decades. Scientists could then turn to the question of the structure of the atoms themselves.

The investigations of electrolysis and spectra gave evidence that matter had an electrical nature and that atoms themselves must be complex systems rather than indivisible entities. The building of models for atoms has revolutionized the physics of the twentieth century. Early attempts involved models made up of particles (electrically negative electrons and the positive nucleus) in electrical interaction with one another. Bohr concluded that the known laws of physics did not apply to atoms, because systems of electrically charged particles could not have the permanence that atoms obviously did have. He therefore assumed new laws, called quantum rules, to supplement the laws of Newton and Maxwell. This approach, while partially successful, soon had to be replaced by a completely new point of view, in which the particle model for matter was abandoned.

Equation 8.1
energy transferred = ΔE
frequency of light = f
Planck's constant = h

$$\Delta E = hf$$

The new concept, introduced by de Broglie and Schrödinger, was to represent the constituents of matter as wave packets instead of as particles. The matter waves are refracted by their interactions with one another, just as particles are deflected by their mutual interactions. Stable atoms are tuned systems in which the electron wave packets oscillate with a characteristic frequency. When the state of the atom changes because radiation is emitted or absorbed, the frequency of the radiation is related to the energy transfer by Bohr's frequency condition (Eq. 8.1).

"The new discoveries made in physics in the last few years, and the ideas and potentialities suggested by them, have had an effect upon the workers in that subject akin to that produced in literature by the Renaissance... In the distance tower still higher peaks, which will yield to those who ascend them still wider prospects, and deepen the feeling, whose truth is emphasized by every advance in science, that `Great are the Works of the Lord."'
J. J. Thomson, 1909

Further developments have led to the recognition that the nucleus of the atom is not indivisible but can undergo spontaneous disintegration in the process called radioactivity. In presently accepted models for the nucleus, protons and neutrons are refracted into a stable standing wave pattern of exceedingly minute physical dimensions by an enormously strong nuclear interaction of very short range. When the state of the nucleus changes, something that happens only during radioactive decay under ordinary conditions on earth, the energy transfer is very much larger than during a change in the electron standing wave pattern in an atom. The technological exploitation of nuclear energy release has led to nuclear reactors for power production as well as to "atomic" (nuclear) bombs.

List of new terms

electrolysis	Planck's constant	gamma rays
electrode	energy level	proton
cathode	ground state	neutron
anode	quantum	atomic number
cathode rays	photoelectric effect	isotope
electron	quantum number	nuclear interaction
electron waves	electron diffraction	nuclear reaction
Thomson's model	wave mechanics	accelerator

alpha rays	valence electron	cyclotron
nucleus	probability	nuclear fission
nuclear model	electron wave packet	chain reaction
planetary model	radioactivity	nuclear reactor
quantum rules	beta rays	control rod

List of symbols

E	energy	$\mathcal{M}$	momentum
ΔE	energy transfer	k	wave number
f	frequency	M_I	inertial mass
h	Planck's constant	v	speed
Δs	size of wave packet (or uncertainty of position)	$\Delta\mathcal{M}$	uncertainty of momentum
		Δk	uncertainty of wave number

Problems

1. Compare the many-interacting-particles models for matter proposed by Greek philosophers, eighteenth-century scientists, and John Dalton. Make reference to: (a) variety in the kinds of particles; (b) interaction among the particles; (c) quantitative relations among the particles.

2. Review and comment upon the evidence that matter has electrically charged constitutents. Which piece of evidence do you find most compelling?

3. Electrons have been described as particles and as waves. Explain briefly what you understand by these terms and how they were appropriate and/or inappropriate.

4. What were some of the problems Bohr was trying to resolve when he introduced his quantum rules? Appraise his success in solving them. (Refer to the Bibliography for additional reading on this subject.)

5. Electron waves are used to study the structure of molecules in a technique called electron diffraction. The inertial mass of electrons is approximately 1×10^{-30} kilogram. Find the wavelength of the electron waves in an electron beam when the electron speed is (a) 1.0×10^6 meters per second; (b) 4.5×10^6 meters per second.

6. Would you expect the diffraction of the electron waves in Problem 5 to give evidence of structure on a scale of sizes equal to, larger than, or smaller than visible light? Explain.

7. George Gamow's *Mr. Tompkins Explores the Atom* (published in 1993 as part of *Mr. Tompkins in Paperback*). describes a world where Planck's constant has a numerical value in the macro domain. Read and comment on this science fiction story.

8. Suppose that Planck's constant had the numerical value of 1 kilogram-(meter)^2 per second, which is a numerical value in the macro domain. Estimate the consequences of this supposition for the behavior of a few macro-domain objects.

9. Estimate the largest possible value of Planck's constant that is compatible with your everyday experience.

10. Prepare a chronology of events in the development of models for atoms and atomic nuclei between 1900 and 1930.

11. Identify one or more explanations or discussions in this chapter that you find inadequate. Describe the general reasons for your judgment (conclusions contradict your ideas, steps in the reasoning have been omitted, words or phrases are meaningless, equations are hard to follow, . . .), and make your criticism as specific as you can.

Bibliography

I. Adler, *Inside the Nucleus,* Day, New York, 1963.

D. L. Anderson, *The Discovery of the Electron,* Van No strand, Princeton, New Jersey, 1964.

J. Bernstein, *Hitler's Uranium Club, The Secret Recordings at Farm Hall,* American Institute of Physics, Woodbury, New York, 1996.

R. T. Beyer, Ed., *Foundations of Nuclear Physics,* Dover Publications, New York, 1949.

D. Bohm, *Causality and Chance in Modern Physics,* Van Nostrand, Princeton, New Jersey, 1957.

M. Born, *The Restless Universe,* Dover Publications, New York, 1957.

R. Calder, *Living With the Atom,* University of Chicago Press, Chicago, Illinois, 1962.

H. Childs, *An American Genius- The Life of Ernest Orlando Lawrence, E.* P. Dutton, New York, 1968.

C. S. Cook, *Structure of Atomic Nuclei,* Van Nostrand, Princeton, New Jersey, 1964.

E. Curie, *Madame Curie,* Doubleday, Garden City, New York, 1937.

W. C. Dampier and M. Dampier, Editors, *Readings in the Literature of Science,* Harper and Row, New York, 1959.

L. de Broglie, *Matter and Light—The New Physics,* Dover Publications, New York, 1939.

L. de Broglie, *The Revolution in Physics,* Noonday Press, New York, 1953.

A. Eddington, *The Nature of the Physical World,* University of Michigan, Ann Arbor, Michigan, 1958.

Editors of *Scientific American, Atomic Power,* Simon and Schuster, New York, 1956.

A. S. Eve, *Rutherford,* Macmillan, New York, 1939.

L. Fermi, *Atoms in the Family: My Life With Enrico Fermi,* University of Chicago Press, Chicago, 1954.

K. W. Ford, *The World of Elementary Particles,* Blaisdell, Waltham, Massachusetts, 1963.

M. Frayn, *Copenhagen* (play), Methuen, London, 1998.

D. H. Frisch and A. M. Thorndike, *Elementary Particles,* Van Nostrand, Princeton, New Jersey, 1964.

G. Gamow, *Atom and Its Nucleus,* Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

G. Gamow, *Mr. Tompkins Explores the Atom,* Cambridge University Press, New York, 1945.

G. Gamow, *Thirty Years That Shook Physics,* Doubleday, Garden City, New York, 1966.

S. Glasstone, *Source Book on Atomic Energy,* Van Nostrand, Princeton, New Jersey, 1958.

B. Greene, *The Elegant Universe—Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory,* Vintage Books, New York, 1999.

O. Hahn, *Otto Hahn, A Scientific Autobiography,* Scribner's, New York, 1966.

S. Hecht, *Explaining the Atom,* Viking, New York, 1954.

P. L. Rose, *Heisenberg and the Nazi Atomic Bomb Project, A Study in German Culture*, University of California Press, Berkeley, 1998.

B. Hoffman, *The Strange Story of the Quantum,* Dover Publications, New York, 1959.

J. Irvine, *The Basis of Modern Physics,* Oliver and Boyd, Edinburgh and London, 1968.

G. Johnson, *Strange Beauty—Murray Gell-Mann and the Revolution in Twentieth-Century Physics*, Vintage Books, New York, 1999.

R. Jungk, *Brighter Than a Thousand Suns,* Harcourt Brace, New York, 1958.

W. B. Mann and S. B. Garfinkel, *Radioactivity and Its Measurement,* Van Nostrand, Princeton, New Jersey, 1966.

A. I. Miller, *Einstein, Picasso—Space, Time and the Beauty That Causes Havoc*, Basic Books, New York, 2001.

R. Moore, *Niels Bohr,* Alfred A. Knopf, New York, 1966.

J. Needham and W. Pagel, *Backgrounds to Modern Science,* Macmillan, New York, 1953.

O. Oldenberg, *Introduction to Atomic Physics,* McGraw-Hill, New York, 1954.

H. Smyth, *Atomic Energy for Military Purposes,* Princeton University Press, Princeton, New Jersey, 1945.

A. K. Solomon, *Why Smash Atoms?,* Penguin Books, Harmondsworth, Middlesex, England, 1959.

J. Tyndall, *Faraday as a Discoverer,* Crowell, New York, 1961.

E. Zimmer, *The Revolution in Physics,* Faber and Faber, London, 1936.

Articles from Scientific American. Some or all of these, plus many others, can be obtained on the Internet at http://www.sciamarchive.org/.

H. Alfven, "Antimatter and Cosmology" (April 1967).

L. Badash, "How the `Newer Alchemy' Was Received" (August 1966).

V. D. Barger and D. B. Cline, "High-Energy Scattering" (December 1967).

H. R. Crane, "The *g* Factor of the Electron" (January 1968).

F. J. Dyson, "Mathematics in the Physical Sciences" (September 1964).

G. Feinberg, "Ordinary Matter" (May 1967).

J. F. Hogerton, "The Arrival of Nuclear Power" (February 1968).

R. B. Leachman, "Nuclear Fission" (August 1965).