DOCUMENT RESUME

ED 046 690 SE 009 209

AUTHOR	Uehling, E. A.
TITLE	Basic Themes of Physics.
INSTITUTION	Commission on Coll. Physics, College Park, Md.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	66
NOTE	62p.; Monograph written for the Conference on the New Instructional Materials in Physics (University of Washington, Seattle, 1965)
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*College Science, *Instructional Materials, *Physics, Reading Materials, Scientific Concepts, Textbooks

ABSTRACT

This monograph constitutes the first five chapters of an introductory course in physics for students who do not intend to become professional physicists. The over-all themes and ideas of physics receive the major emphasis, with little stress placed on mathematical manipulation. The style of presentation is mainly a combination of the axiomatic and the historical. This mode of writing, as well as organization and selection of topics, was based on the author's estimation of what would most enhance student interest and understanding. The five chapters in order of presentation are Introduction, Kinematical Preliminaries, Dynamical Principles, Force, and Motion Under the Influence of Gravitational Forces. (PR)

Basic Themes of Physics

E. A. UEHLING

University of Washington

GENERAL PREFACE

This monograph was written for the Conference on the New Instructional Materials in Physics, held at the University of Washington in the summer of 1965. The general purpose of the conference was to create effective ways of presenting physics to college students who are not preparing to become professional physicists. Such an audience might include prospective secondary school physics teachers, prospective practitioners of other sciences, and those who wish to learn physics as one component of a liberal education.

At the Conference some 40 physicists and 12 filmmakers and designers worked for periods ranging from four to nine weeks. The central task, certainly the one in which most physicists participated, was the writing of monographs.

Although there was no consensus on a single approach, many writers felt that their presentations ought to put more than the customary emphasis on physical insight and synthesis. Moreover, the treatment was to be "multi-level" --- that is, each monograph would consist of several sections arranged in increasing order of sophistication. Such papers, it was hoped, could be readily introduced into existing courses or provide the basis for new kinds of courses.

Monographs were written in four content areas: Forces and Fields, Quantum Mechanics, Thermal and Statistical Physics, and the Structure and Properties of Matter. Topic selections and general outlines were only loosely coordinated within each area in order to leave authors free to invent new approaches. In point of fact, however, a number of monographs do relate to others in complementary ways, a result of their authors' close, informal interaction.

Because of stringent time limitations, few of the monographs have been completed, and none has been extensively rewritten. Indeed, most writers feel that they are barely more than clean first drafts. Yet, because of the highly experimental nature of the undertaking, it is essential that these manuscripts be made available for careful review by other physicists and for trial use with students. Much effort, therefore, has gone into publishing them in a readable format intended to facilitate serious consideration.

So many people have contributed to the project that complete acknowledgement is not possible. The National Science Foundation supported the Conference. The staff of the Commission on College Physics, led by E. Leonard Jossem, and that of the University of Washington physics department, led by Ronald Geballe and Ernest M. Henley, carried the heavy burden of organization. Walter C. Michels, Lyman G. Parratt, and George M. Volkoff read and criticized manuscripts at a critical stage in the writing. Judith Bregman, Edward Gerjuoy, Ernest M. Henley, and Lawrence Wilets read manuscripts editorially. Martha Ellis and Margery Lang did the technical editing; Ann Widditsch supervised the initial typing and assembled the final drafts. James Grunbaum designed the format and, assisted in Seattle by Roselyn Pape, directed the art preparation. Richard A. Mould has helped in all phases of readying manuscripts for the printer. Finally, and crucially, Jay F. Wilson, of the D. Van Nostrand Company, served as Managing Editor. For the hard work and steadfast support of all these persons and many others, I am deeply grateful.

Edward D. Lambe
Chairman, Panel on the
New Instructional Materials
Commission on College Physics

BASIC THEMES OF PHYSICS

PREFACE

These pages constitute the opening chapters of an introduction to physics for students who do not intend necessarily to make physics their lifetime major interest. The emphasis is deliberately placed on the over-all themes and general ideas of physics and very little on the manipulative processes. For this reason the textual material tends to be discursive, but at the same time an effort is made to be concise. In no sense is the student being talked down to. On the contrary he should feel that with every new topic presented his mind is being challenged and that a rather considerable intellectual effort on his part will be required to fully comprehend what is being said and implied. Since mathematical prerequisites are not stressed, the student will recognize that what difficulties there are in the understanding of physics are not in the mathematics which is so conveniently used in the discussion of physical ideas but in the ideas themselves.

The organization and selection of topics are based on one view of what is most likely to stimulate the students' interest and enthusiasm for physics and what will at the same time contribute to a genuine understanding. The point of view taken here is that in order to achieve these objectives, physics may well be regarded as a study of the elementary processes of nature, even at the beginning level. The principle theme of the course, at least in its early stages, is motion and the manner in which systems evolve in time. The course begins in the way other courses do with a study of kinematics, but in a somewhat different context as determined by the tenor of the Introduction as well as by the discursive passages accompanying the presentation of each new idea. Also, important differences in comparison with standard methods tend to accumulate as the subject is developed. One of the most important is the early discussion of force in the context of basic interactions, the presumption being that the number of types of interactions in nature is small and knowable. The status of our current knowledge in this respect is stated immediately following the discussion of Newton's laws of motion. Other differences that occur early in the text are the systematic discussion of the three important simple categories of motion in the section on kinematics before any dynamics is mentioned, the attempt to enliven the subject of kinematics by calling upon the students' recently acquired understanding of acceleration in circular motion and his knowledge of free-fall acceleration to predict the velocity of an earth satellite, and the introduction of relativity and its use under a Galilean transformation to establish the principle of conservation of momentum. From this point on the organization of material is not wholly determined at this time, but a long chapter on electricity and magnetism, omitting most of the historical introduction to electrostatics, will come early; there will be a section on the molecular-atomic constitution of matter with kinetic theory and thermodynamics in that order; and there will be a section on relativity and light. This will be followed by quantum theory, the structure of atoms and nuclei, and some elementary particle physics.

The mode of presentation throughout the text is a combination of the axiomatic and historical, with the balance chosen hopefully in such a way that the students' interest is the more strongly enhanced. Consequently, whenever it seems relevant to discuss the human aspects of progress in science and something of the manner in which scientific ideas originate and are developed, historical aspects have been stressed. At the same time it is assumed that students live in a modern world in which the notion of atoms and molecules, atomic energy and satellites, is commonplace, and not all of the historical development needs to be belabored. In a few sections the presentation is strictly axiomatic.

The first five sections which follow constitute the bare text for those portions of the course and may be regarded as typical of what the bare text for the remainder will be like. This text must eventually be enlivened with more figures and with examples and problems. A course along lines similar to this has been given once, and another attempt has been in progress during the summer of 1965. Consequently there is an associated list of exercises, problems, quizzes, and examinations, but this material is not included at this time.

Edwin A. Uehling

CONTENTS

1 INTRODUCTION

Physics deals essentially with the elementary processes of nature. Its method is to isolate such processes conceptually, and if possible, experimentally, and thus to study the various processes occuring in nature under the simplest possible conditions. The elements of such a study are the physical entities composing the system, the properties which these entities are observed to possess, the nature of the interaction between them, and the general principles which govern the way in which the over-all system evolves in time. In order to obtain a description in these terms the appropriate concepts must be developed and the principles of physical behavior must be discovered.

As an example of an elementary process occurring in our everyday experience, the free fall of an object near the earth's surface may be mentioned. That an object falls when released is known from earliest childhood. But to analyze this motion as Galileo was the first to do almost four centuries ago and to discover that the motion which takes place is of a particularly simple kind required something more than mere observation. Among other things an abstraction from the actual situation occurring in nature, thus eliminating certain extraneous elements such as the resistance of the air, was required. Also, a rather considerable exercise of the imagination was needed in order to devise experiments and to establish tests leading eventually to an unambiguous characterization of the particular motion under consideration. But the student will also note something more. In even so simple a physical situation as this there are many problems. We have spoken of only one of them: the kind of motion. Even to answer this question one has to know something about motion in a general sense, what it is, what the concepts are in terms of which it is described. These are some of the kinds of questions people were able to answer up to the time of Galileo and to which he himself made important contributions. But having answered these questions, one finds immediately that there are others, questions which in many cases are suggested by the answers already given. In the case of free fall one would naturally ask, after having learned that the motion is of a certain kind, why this kind of motion occurs in this particular physical situation in contrast with other kinds of motion which occur in other physical situations. In effect, what is the essence of motion, and what is it that determines whether it is of one special kind or another.

As a second example from our everyday experience, let us mention the common observation that there are fringes of color which are seen by the most casual observer under a great variety of circumstances: the so-called irridence of a thin film of oil on water, reflected light from small bits of glass under certain circumstances, light shining through the mist at a waterfall, and on a larger scale, the rainbow itself. Camera enthusiasts are aware also of an analogous effect called chromatic aberration, a property of all simple lenses. The physical situation common to all these phenomena is the passage of ordinary light, i.e., the light of the sun, through transparent materials such as oil, water and glass. Clearly, the physical situation can be idealized, and quite clearly also some idealization is required before an interpretation free of irrelevancies can be obtained. The first one to do this was Newton. His immediate problem was the elimination of color fringes surrounding the image produced by a lens.

He soon realized that the problem had to be reduced to simpler terms, and he began the study of the properties of white light using prisms. In this way he showed that ordinary light is composed of many colors and that a prism has the property of decomposing light into its various color components. This information may be shown to be sufficient to explain the appearance of color fringes around the image produced by a lens. But as in our previous example, it raises new questions, and enhances the interest in old questions: in this case, the nature of light itself, and then, the nature of the interaction between light and the transmitting medium. The physical situation is now not as simple as in the case of free fall. We will find that in the latter the appropriate abstraction is to consider simply two objects, the earth and the falling body, attracting each other with a force of a certain strength which we will be able to determine and describe. In the former we are dealing with the interaction of light and matter. Obviously, we will have to know a great deal about light before we can even begin the discussion of this problem. Also we will have to know and understand the properties of atoms, the ultimate constituents of the matter through which the light is passing. And finally, we will have to understand how light interacts with atoms. Thus, a further abstraction in this problem will be the description of how light interacts with a single atom.

Much of modern physics is far removed from the realm of everyday observation. The frontier of physics still spans the entire universe from the cosmological to the subatomic, but the greatest activity and interest is naturally at the two extremes of this range. There is one important difference compared with the old physics. As we have just seen in two typical examples, the study of elementary processes in the old physics was a natural consequence of the questions people ask concerning the natural phenomena of everyday experience. By a process of isolation and idealization the actual phenomenon is decomposed into its parts and freed from effects which may be regarded as temporarily irrelevant. This leads eventually to a decomposition of a large physical problem into a number of smaller ones, some of which may be regarded as descriptive of specific elementary processes in nature. The elementary processes of modern physics are of a different character. Many of these processes are discovered simply by pushing the modern techniques of observation to their limits. The techniques are so powerful that completely unsuspected phenomena are frequently brought to light. Experiments performed for a different purpose will lead in this way to results of the most astonishing kind. Thus, a new elementary process, one which is not merely a simple component in a more complex physical situation, is disclosed. Since the process was unsuspected, the reason for its existence in nature is not known, and frequently cannot be surmised. Thus the study of the purpose in nature of elementary processes and the associated elementary particles discovered by the methods of modern physics becomes one of the larger problems of modern physics. At the present time there are many such cases. A typical example is the muon, an elementary particle of modern physics which, insofar as we have been able to determine after a quarter century of study is identical with the electron, only heavier. The muon was discovered when people looked for a heavy particle, which, according to theory, would account for a strong interaction force between two other elementary particles, the proton and neutron. It was another ten years before this latter particle, called the pion, was discovered. In the meantime the muon, having properties like the electron, was not capable of providing the desired interaction force, and subsequently, and up to this day, no purpose in nature for this particle has been found.

Finally, we will note that the elementary processes of nature with which physics must deal are not necessarily confined to what is generally regarded as the physical universe in contrast to the biological. The biologists are finding that many, perhaps all, of the elementary biological processes are physical (or chemical) in character. As an example we will note the remarkable union of biology and physics which seems to be developing in connection with the efforts of biologists and geneticists to understand the hereditary mechanism. Biological experiments of great subtlety have proven that the gene, the locus in the chromosome of specific biological properties, is essentially permanent; i.e., it is transmitted from generation to generation and through repeated cell divisions and duplications without any change whatever. A possible explanation of this permanence of structure is believed to lie in the quantum theory. Furthermore, the genetic code contained in the DNA molecule and involving the linkage of specific molecular groups each containing a relatively small number of atoms must be physical in character. Modern physical theory based on quantum concepts has been developing steadily ever since the first hints of quantum phenomena were discovered by Planck and Einstein in 1900 and 1905. Modern genetics and genetic theory, starting with de Vries, Correns, and Tschermak, has also been developing since the beginning of this century. The possible union of these two theories at this time when both have reached a certain maturity is a matter of greatest interest and importance to both the biologists and physicists, and of course to people in general.

These examples may provide the beginning student with some idea of what physics is about before starting the course. But examples alone cannot give a very clear picture, either of the vast scope of the subject, or of the depth and beauty of its various parts. Not even a course in physics can do that adequately. But by a selection of the material, and by a choice of emphasis, we can perhaps hope to convey something of the spirit of physics as well as something of the physical content. Our objective should be to provide some understanding of what is basic and exciting in modern physics, some conception of science in general as a continuous state of intellectual inquiry, some notion of what is contained in the large body of knowledge that is the result of that inquiry, and some conception of the ever expanding frontier which separates that body of knowledge from that which is presently unknown.

In order to accomplish these objectives the principal ideas of physics as they came to be understood in the course of man's study of natural phenomena will be presented. The emphasis throughout will be on ideas as described in terms of concepts, general principles, and the connections between principles and phenomena, rather than on detailed applications. Thus, it will not be expected that the student will acquire facility in the application of the stated principles to new situations, but it will be expected that the student will be able to describe selected physical phenomena in terms of general principles, and to this extent at least, achieve an understanding of basic principles and some of their more immediate implications.

Since the emphasis will be on the ideas of physics rather than on strict analysis, much of the discussion and some of the description of actual physical situations can be made in verbal terms; i.e., without the use of mathematical symbolism. However, since all logical processes are in essence mathematical, the student will find that avoidance of symbolism is little more than a subterfuge. Consequently, the student will find it helpful to recall what he can from previous courses in algebra and geometry, and to acquire a certain facility in the use of algebraic and geometric methods. Early in this course he or she will be assisted in doing this by a review of what is needed for present purposes. There will also be some instruction in the concepts and definitions of trigonometry, in the use of coordinate systems, in the meaning and graphing of simple algebraic functions, and in the nature of scalars and vectors. Also we will describe and use the method of analysis which is based on making small changes in each of two variables which depend on each other, comparing them, and studying their ratio in the limit that the changes are very small. The resulting concept is a basic one in the calculus, and is called a rate of change.

2 KINEMATICAL PRELIMINARIES

We begin the study of physics by putting our attention on a single property of the motion of an object: its location in space at various times. This is the kind of description we give when we ask, for example, how far down a certain road an automobile may have traveled after leaving its starting point when moving at a specified speed, say 30 miles per hour. Similarly, we may ask about the distance of travel of a baseball along its curved path after leaving the pitcher's hand, and again, the question can be answered if we know the speed with which the ball was pitched and can calculate the changing speed of the ball as it moves. Note that when we ask such questions about the automobile or the baseball we are abstracting from certain other properties of these objects. The baseball, for example, may be rotating at the same time that it is traveling forward. Clearly, we are ignoring this other motion, and closer examination will disclose that it is legitimate under certain circumstances to do so. In the same manner we are ignoring also numerous other physical properties of the object under discussion in our attempt to describe its forward motion. This process of abstraction in physics is one which must be justified as we go along. It is at the same time an essential process. No physical situation can ever be described in all of its details. We are forced for the sake of clarity of argument, as well as by the sheer complexity of natural phenomena, to reduce each physical situation to just those relevant elements in which we happen to be interested. This is what is meant by abstraction in physics. We would make no progress without it.

2.1 POSITION

We consider now the motion of an object as a whole. We speak of this kind of motion as translational motion. The object is being translated from one point in space to another. Since we are ignoring all other properties, the object itself may be conceived as something which in fact possesses no other properties than those which are relevant to its behavior in translational motion. Since, in effect, we will have selected a single point in the object to which we consider position locating measurements to have been made, e.g., the front of the automobile in its travel along the road, it is only this selected point of the object which is of immediate interest to us. Consequently, we ignore everything else and put our attention on the selected point. This leads us to the concept of the point particle. We will presently attribute other properties to the point particle than mere location in space, but we will legitimately abstract at least from all properties of extension possessed by the object in question. Clearly, this is a tremendous idealization when we are speaking of an object as large and complicated in structure as an automobile, or even a baseball, which have dimensions measured in feet or inches. It will seem like much less of an idealization when the object under consideration is a proton or electron whose dimensions are very much less; e.g., radius of the order 3×10^{-13} cm for the electron.

The problem of translational motion is now reduced to the specification of the location of a point particle at various instants of time. We will here consider the first part of this statement; i.e., the specification of position. The way in which we do this depends on the situation. Let us consider several cases in turn.

Suppose that a point particle is constrained to move along a straight line; e.g., the automobile on a straight road. Regard the line as extending indefinitely in both directions. In order to locate the object along this indefinitely long straight line we must have a reference point. Once the reference point has been selected, all positions along the line can be specified in terms of the distance to the reference point. We note, however, that the object may be either on one side or the other of the reference point. For definiteness we call distance in one direction positive and in the other negative. We find it convenient to denote the distances by a symbol. For example, if the line is a horizontal one and we measure distances to the left and right from a certain reference point, a common convention would be to denote the distance by a letter x; if the object is at a distance x_1 to the right, the position will be denoted by x_1 where x_1 is a positive number; if it is at a distance x_2 to the left, the position will be denoted by x_2 where x_2 is a negative number. The reference point is called the origin.

A similar convention may be used for motion along lines which are not straight; e.g., the motion of an object at the end of a string in a circle. Since the point particle is constrained in its motion so that it must always lie on some given line, straight or curved, its position is uniquely determined by a single number, say s in this case, where s is measured from an arbitrarily chosen fixed reference point on the line, and is positive or negative according to convention, depending on which side of the reference point it may be located.

The position of a point particle in a plane may be specified by a simple extension of these procedures. One now requires two reference lines which may be chosen arbitrarily in a variety of ways. The simplest choice is to use two straight lines of infinite extent which intersect at right angles. Denote the point of intersection as the reference point. We call it the origin of rectangular coordinates. Denote distances along one line by the symbol x which is positive or negative according to which side of the origin a given point may lie, and denote distances along the other line by a symbol y which is similarly positive or negative according to which side of the origin a given point may lie along the y line. The point particle under consideration is now regarded as moving in a plane, and we take the plane defined by the xy coordinate system to be coincident with the plane in which the particle is constrained to move. In order to locate the point particle we drop perpendiculars from it to the x and y axes which intersect at say $x = x_1$ and $y = y_1$ where x_1 and y_1 constitute a pair of numbers. These numbers uniquely define the position of the particle.

The extension of these methods to three-dimensional space is obvious. We now use three straight lines which are mutually perpendicular and intersect at a single point, the common origin of the three-dimensional rectangular coordinate system (the so-called system of Cartesian coordinates). We call the three lines the x, y, and z axes of the coordinate system and we measure distances along these lines from the origin and denote the results by the same letters x, y, and z. The location of a point particle in space is now uniquely specified by the three numbers x_1, y_1, and z_1, which are the numerical values of the intersections on the x, y, and z axes, respectively, of perpendiculars drawn from the position of the particle to the three coordinate axes.

2.2 VELOCITY

If the point particle is in motion its position changes with time. We may begin again by considering motion along a straight line. We observe immediately that there are two distinct cases. The first of these is such that the distances traveled in equal intervals of time are the same, no matter when the observation is made. If we divide any interval of distance traveled by the time it took to travel that distance we obtain the same result, independent of which intervals are chosen. In symbols,

$$v = \frac{x_2 - x_1}{t_2 - t_1},$$

where $x_2 - x_1$ is the interval of distance considered, and $t_2 - t_1$ is the corresponding time interval. The ratio, denoted by the symbol v, is called the velocity of motion, or simply the velocity. For the case considered v is a constant. Frequently we will simplify the notation by writing

$$v = \frac{x}{t},$$

which has precisely the same meaning as the previous relation. The only difference is that the zero values of x and t are in effect chosen at the beginning of the motion. We observe that the case of constant velocity can be equally well characterized as the case of uniformly increasing distance with time; i.e., we may write the relation in the form

$$x = vt,$$

which shows that x increases proportionally with the increase in time.

All other cases fall into the second category; i.e., the motion is not at constant velocity. We first observe that if the velocity is not a constant, the ratio of $x_2 - x_1$ and $t_2 - t_1$, which was used previously to define velocity, will now have a value which depends on which intervals of x and t are chosen. Also it depends on the size of the intervals. These facts suggest that the only sensible definition of velocity which we can make is one which makes use of position and time intervals which are arbitrarily small. Here and elsewhere we will denote arbitrarily small intervals by placing a symbol Δ before the quantity we are considering. Thus we will denote an arbitrarily small time interval by the symbol Δt. Let Δx be the small distance traveled in the small interval of time Δt. Then using the same definition of velocity as was used in the case of constant velocity we define

$$v = \frac{\Delta x}{\Delta t}.$$

This definition is still not quite satisfactory however. A little consideration shows that v still depends on the size of Δx and the associated Δt but that this dependence tends to disappear as we make the intervals smaller and smaller. Thus, we consider that the ratio is to be evaluated in the limit that both Δx and Δt are reduced to zero. We express the definition formally by writing

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}.$$

The velocity so defined is called the instantaneous velocity. It is a definition that can be used under all circumstances. Further clarification of its meaning may be obtained from a graphical construction. Suppose we plot x as a function of t in a given case. Examination of such a plot will show that the v obtained from the definition as given here is simply the slope of the curve of x as a function of t. Both the slope and the velocity have well-defined instantaneous values at arbitrarily selected values of t or x.

The extension of these definitions to motion in a plane or in three-dimensional space is most easily made if we consider that the motion parallel to the different coordinate axes of a rectangular coordinate system are independent of each other. Thus, in two dimensions we have motion at velocity v_x parallel to the x axis and at velocity v_y parallel to the y axis. The velocities v_x and v_y are called the components of the actual velocity v in the plane. Their definitions are given separately in precisely the same way as for one-dimensional motion. Thus

$$v_x = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$$

$$v_y = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t}.$$

The extension to motion in three dimensions requires nothing more than the addition of the equation for the third component, i.e.,

$$v_z = \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t}.$$

It is convenient as well as instructive to speak of velocity as the rate of change of position. It is our first example of a physical variable which is defined as the rate of change of another. Quite clearly, zero velocity corresponds to zero rate of change of position (no motion), low velocity corresponds to position changes which are taking place at a low rate, and so on. The symbolic definitions given above are simply a way of expressing the idea of rate of change precisely and in a form suitable for quantitative use.

2.3 ACCELERATION

In the same way that velocity is defined as rate of change of position, acceleration is defined as rate of change of velocity. A few examples will help to clarify this concept. Motion at constant velocity is said to be motion at zero acceleration, i.e., the rate of change of velocity with time is zero. The next simplest case is that in which the velocity increases (or decreases) uniformly with time. By this we mean that the velocity is changing, and that the amount of change is the same in each succeeding second. This is motion at constant acceleration. Consider for example motion along a straight line. If the velocity is changing uniformly with the time its value v at any time t must be given by

$$v = v_0 + at,$$

where v_0 is the velocity at $t = 0$, a is a constant, and v increases or decreases uniformly with t depending on whether a is positive or negative. We call a the acceleration (constant in this case). By solving for the acceleration and expressing the result in the form

$$a = \frac{v - v_0}{t},$$

we observe that a is the rate of change of velocity; i.e., the change of velocity per second. A more general form of the definition is

$$a = \frac{\Delta v}{\Delta t},$$

where Δv is the change of velocity in the time Δt, and, for the case of constant acceleration, the ratio is a constant independent of the size of the intervals or of the way in which they are chosen. The analogy with the definition of constant velocity as the ratio of Δx to Δt is apparent.

If the acceleration is not a constant we must proceed in the same way as for the case of velocity which is not constant. As for the latter, we are able to define the instantaneous values by choosing arbitrarily small intervals and taking the limit as the intervals go to zero. Thus, the general definition of acceleration for one-dimensional motion is

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}.$$

The extension to two- and three-dimensional motion is the same as for the velocity. Also it is instructive to plot v as a function of t and to observe that the instantaneous acceleration as here defined is the same as the slope of the curve at corresponding values of t. This is analogous to the correspondence between instantaneous velocity and the slope of curve in the plot of x as a function of t.

Position, velocity, and acceleration constitute the complete set of kinematical quantities required in order to give a complete description of any motion. The reason that we require these three and no more will become apparent later.

We will conclude with some remarks about the units in which to measure the kinematical quantities. We will most frequently consider distance to be measured in centimeters (cm) and time to be measured in seconds (sec). Then the proper units for position, velocity, and acceleration, as may be verified by consulting the definitions, are cm, cm/sec, and cm/sec^2, respectively.

2.4 CLASSIFICATION OF MOTION

The first successful characterization of any natural motion was made by Galileo after a series of ingenious experiments at the University of Pisa between the years 1590 and 1600. These experiments demonstrated that all motion at the surface of the earth, due to the influence of the earth's gravitational field, and abstracting from the effects of friction and other influences, is motion at constant acceleration. It is probably this result together with the observations of Kepler on the motions of the planets which stimulated Newton and led to his formulation of the equations of motion less than a century later. Before Newton, people did not know the basis upon which a classification of motion was possible, a fact which renders the achievement of Galileo all the more remarkable. Since the classification will be helpful in our subsequent discussions we will anticipate the basis for the classification, deferring its justification until after we introduce the equations of motion. The basis is simply the nature of the acceleration. We will now describe the kinematical properties of several important types of motion classified on this basis.

2.4.1 Motion At Constant Acceleration

We will restrict the present considerations to motion along a straight line and take the acceleration a to be directed along this line in the positive sense if a is positive. Then, as we have already seen, the velocity at any time t is given by

$$v = v_0 + at,$$

where v_0 is the velocity (positive or negative) at t = 0. We are interested now in calculating the distance of travel during a time t if at t = 0 the body is at x = 0. This is easily calculated by noting that for this case of uniformly increasing velocity the average velocity $\overline{v}$ during the time t is equal to half the sum of the initial and final velocities; i.e.,

$$\overline{v} = \tfrac{1}{2}(v + v_0).$$

Then the distance traveled is

$$x = \overline{v}t = \tfrac{1}{2}(v + v_0)t$$
$$= v_0 t + \tfrac{1}{2}at^2.$$

This is the characteristic result for motion at constant acceleration; the distance through which the object moves in a time t increases as the square of the time. As noted by Galileo, one can easily show from this result that the distances through which the object moves in successive unit intervals of time are in the ratio of the odd integers.

Also, as shown by Galileo on the basis of actual experiment, these are the relations which correctly describe motion in the earth's gravitational field (bodies rolling down inclined planes and bodies falling freely toward the surface of the earth). Galileo's stop watches (water clocks) were not sufficiently accurate to provide him with reliable data on freely falling bodies, but he was able to establish beyond doubt that the motion with which he was dealing was one of constant acceleration. Consequently, one of our first important results is that motion at the surface of the earth takes place at constant acceleration. Although Galileo could not measure the magnitude of the acceleration very accurately, its value is in fact easily determined, and methods for its determination will be discussed in class. Anticipating the measurements, we will now give the result. We denote the value of the acceleration of gravity at the earth's surface by the letter g in order to set it apart from all other accelerations which we will continue to denote by the letter a. The result of the measurement is $g = 980$ cm/sec^2 approximately and it is directed very nearly toward the center of the earth. For reasons to be discussed at a later time, its direction (with respect to a line drawn toward the center of the earth), and its magnitude vary slightly over the surface of the earth.

Galileo's contribution to the understanding of the principles of motion is so great, and his thought processes (as demonstrated by the kinds of experiments which he performed and by the logic of his arguments), are so beautiful, that every student of science, casual or not, should want to make some further study of his work. An excellent summary together with excerpts from the original writings is to be found in Shamos, Great Experiments in Physics, pp. 13-35.[1]

One additional relation valid for straight-line motion at constant acceleration is of interest. As we have seen the defining equation for acceleration automatically gives the velocity as a function of the time in the special case of constant acceleration. The question may be asked whether or not we can give an expression for the velocity as a function of distance. A little consideration will show that all we need to do in order to obtain such a relation is to eliminate the time between two relations involving v and t and x and t. One way of doing this is as follows: Start with the two relations

$$v = v_0 + at$$
$$x = \bar{v}t = \tfrac{1}{2}(v + v_0)t;$$

rewrite them in the form

$$v - v_0 = at$$
$$v + v_0 = 2x/t;$$

now multiply the two equations by each other to give

$$v^2 - v_0^2 = 2ax.$$

At this stage the student will find his understanding of what has been accomplished enhanced by making applications to a few specific situations. It will be helpful also to plot some of the relations. Consider for example the simple straight-line plot of v as a function of t for each of the four cases obtained by taking combinations of positive and negative a with positive and negative v_0, then in each case consider the behavior of x as a function of t and v as a function of x. An important special case is that of free fall along a vertical line in the earth's field. Consider

[1]Morris H. Shamos, (ed.), Great Experiments in Physics (Holt, Rinehart and Winston, Inc., New York, 1960).

problems in which the object is given initial velocities upward and downward, and use various sign conventions in solving the problem.

2.4.2 Motion In A Vertical Plane At the Earth's Surface

On the basis of the discussion given in the preceding section, we already know how to characterize this type of motion. We have learned that objects moving under the influence of the earth's gravitational field experience an acceleration directed toward the center of the earth equal to $g = 980$ cm/sec.^{2}. In addition we have already stated that motion in a plane can be decomposed into motions along an x axis and a mutually perpendicular y axis. Let us now take the y axis to be the vertical axis. Then the y motion is at constant acceleration g directed downward. Since the y axis is vertical, the x axis is horizontal. There is no acceleration along this axis. Consequently, the x motion is at constant velocity.

Before preceeding further we must say something more about the nature of the problem to be discussed. Also we must make a convention in regard to signs. The motions we are about to consider are those which are initiated by projecting an object with some given initial velocity in some given direction and then allowing the object to move freely under the influence only of the earth's gravitational field. A convenient way of specifying the initial conditions is to give the two components of initial velocity along the x and y axes. We denote these two components as v_{0x} and v_{0y} respectively. Next we must adopt a convention with respect to signs. For the x axis it is convenient to take the positive direction as the direction of v_{0x}. Then v_{0x} enters the equations as a positive quantity. For the y axis we may take the positive direction to be upward (in which case the acceleration is $-g$ and v_{0y} is positive if directed upward), or we may take it to be downward (in which case the acceleration is $+g$ and v_{0y} is negative if directed upward). Let us here choose the positive direction of the y axis to be upward. Then using the results already derived for distance as a function of the time in constantly accelerated motion we write for both the x and y motions the relations

$$x = v_{0x}t$$
$$y = v_{0y}t - \tfrac{1}{2} gt^2.$$

According to these equations the object is definitely located in position at all later times t if, in the same arbitrarily chosen coordinate system, the object was at the origin at the time $t = 0$.

It is of interest to consider the nature of the path (the trajectory), followed by the object in its free fall. The equation for this path may be obtained by eliminating t between the two equations. This elimination gives

$$y = \frac{v_{0y}}{v_{0x}} x - \frac{1}{2} \frac{g}{v_{0x}^2} x^2.$$

The characteristic feature of this equation is that y varies as the square of x in addition to having a contribution proportional to it. This feature gives the curve a characteristic shape. Paths in a plane which possess this feature of one variable varying as the square of the other are called parabolas. Thus, the trajectories of freely falling bodies are parabolas. The student will find it instructive to plot a few of these curves using several different choices of v_{0x} and v_{0y}.

The student may at this point raise an important question. If he recalls the theorem of Pythagorus he knows that the actual displacement of the object in the time t (the straight-line distance between the starting point and a selected point on the trajectory), is given by

$$s = \sqrt{x^2 + y^2}.$$

One may well ask what is the actual velocity at a selected point on the trajectory. To begin with we know the components. They are

$$v_x = v_{0x}$$
$$v_y = v_{0y} - gt,$$

where in the second relation we are simply using again the definition of constant acceleration and recalling our convention in regard to signs. We can now show that the actual velocity v is obtained from the components v_x and v_y in the same way that the actual displacement s is obtained from the component distances x and y. To show this consider a small increment of displacement Δs along the path with the components Δx and Δy. Then by the Pythagorean theorem

$$\Delta s^2 = \Delta x^2 + \Delta y^2 .$$

Dividing by Δt^2 we obtain

$$\left(\frac{\Delta s}{\Delta t}\right)^2 = \left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta v}{\Delta t}\right)^2 .$$

But by definition of velocity this is simply

$$v^2 = v_x{}^2 + v_y{}^2 .$$

Quantities such as s (with components x, y) and v (with components v_x, v_y) are called vectors. We denote them as $\vec{s}$ and $\vec{v}$. They possess the important property of composition by addition of the squares of the rectangular components. They are quantities possessing both magnitude and direction. They may be completely expressed, as we have done, in terms of their components. Or they may be expressed in other ways. One other way which we will find convenient is to give the magnitude of the vector and the angle which the vector makes with some specified direction. Consider for example the velocity on the parabolic trajectory which we have been considering. The magnitude of this velocity is

$$v = \sqrt{v_x{}^2 + v_y{}^2} .$$

Now we must also give its direction. One way is to specify the angle θ between the direction of the vector and the x (or y) axis. Suppose we use the x axis. Then in terms of the components we find it easy to show that $\tan \theta = v_y/v_x$. The two specifications ($\vec{v}$ in terms of v_x and v_y, or $\vec{v}$ in terms of v and θ) are completely equivalent. We use whatever is most convenient.

Perhaps we should add one additional point in regard to our whole procedure. We have assumed without proof of any kind that all motion can be decomposed into mutually perpendicular motions which behave independently, and that these motions may then be compounded to give the resultant motion. The only proof of the validity of this procedure is that which is obtained from experiment. Independence of motion has been assumed and a parabolic trajectory in agreement with observation has been predicted. It is this interplay of intuition, hypothesis, experiment, and comparison with the predictions of theory that provides the basis for the acceptance and rejection of ideas and the gradual accumulation of what we classify as scientific knowledge. Galileo was among the first to recognize this. The motion we have been considering was in fact completely analyzed by him, and in much the same way as we have done it.

The motion we have been describing is often called projectile motion because of its obvious applications. For a further discussion of this motion see, for example, Holton, Introduction to Physical Science, pp. 36-53; Orear, Fundamental Physics, pp. 26-31; and, Shortley and Williams, Elements of Physics, pp. 77-79.[2]

[2] Gerald Holton, Introduction to Concepts and Theories in Physical Science (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1952); Jay Orear, Fundamental Physics (John Wiley & Sons, Inc., New York, 1961); George Shortley and Dudley Williams, Elements of Physics (Prentice-Hall, Inc., New York, 1953).

2.4.3 Circular Motion

A very common type of natural motion is motion at constant speed in a circle. For example, many of the planets move in orbits which approximate very closely to a circle, and the speeds are nearly constant. The motion of electrons in atomic orbits can be described quite satisfactorily for many purposes by using a model based on circular orbits and constant speeds. Also in our everyday experience we encounter such motions frequently, e.g., an automobile moving at constant speed around a curve of constant radius, the rider in a merry-go-round, the contents of a cream separator or other centrifuge. Thus, there are many examples and we must conclude that this is an important type of motion.

We begin our consideration of this motion by noting that our statements in regard to it are of a different character than in the two previous cases. Instead of specifying the acceleration and then asking for the type of motion, as we did previously, we are now specifying the type of motion, and we shall take as our problem the specification of the acceleration. From the point of view of a classification of motions the actual starting point in any given case is of course unimportant.

That there is an acceleration must be obvious from what has already been said. We have defined acceleration as the rate of change of velocity. In symbols for straight-line motion

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}.$$

But the present case is not one involving straight-line motion. Velocity is a vector, and though its magnitude is a constant (uniform speed around the circle), its direction is changing continuously. Consequently the velocity is in fact changing and there must be an acceleration. The acceleration is also a vector. Thus, it has both magnitude and direction, and it is these properties which we now wish to determine.

The analysis is most easily given in terms of a graphical construction. The procedure is a simple one and will be carried out in class. It is described also in many texts and students will want to refer to these other treatments. See, for example, Holton, Introduction of Physical Science, pp. 93-94, and Orear, Fundamental Physics, pp. 31-34 (see footnote 2).

The result of the analysis is as follows: Any object moving at constant speed v in a circle of radius r is being continuously accelerated toward the center of the circle. The magnitude of the acceleration has the constant value

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

Note carefully both parts of this statement: The acceleration has a constant magnitude v^2/r and a constantly changing direction, namely, the direction is always toward the center of the circle. The direction of the acceleration experienced by the moving object is perpendicular to its direction of motion.

We will later consider a number of interesting applications of this result. For the present let us consider an application which we can make easily to the motion of an object around the periphery of the earth (an earth satellite). Such an object is unsupported and consequently it falls in the earth's gravitational field with an acceleration $g = 980$ cm/sec^2, if it is not very high above the earth's surface and the value of g is about the same as at the earth's surface. But now we consider the object to have a forward velocity v. At low velocities it will not move very far over the surface of the earth before it hits the ground. The trajectory will be one of the parabolic trajectories already described. But suppose we increase v to higher and higher values. Then the point of contact with the earth will move forward, and eventually it will be sufficiently far forward that we can no longer regard the object as moving in a simple x, y coordinate system with acceleration g always directed along the y axis. In other words, we must take into account the fact that the earth's surface is in effect falling away from the object. When this happens at a rate such that the object in its fall at constant accleration g remains at a fixed distance above the earth's surface, the object will be moving at constant speed v in a circle of radius equal to that of the earth or a little larger. This is one of the possible orbits of an earth satellite. Setting the acceleration for motion in a circle equal to the free-fall acceleration we have

$$g = \frac{v^2}{R_e} ,$$

where R_e is the radius of the earth. Using $R_e = 6.36 \times 10^8$ cm one finds

$$v = \sqrt{gR_e} = 7.9 \times 10^5 \text{ cm/sec}$$
$$= 18000 \text{ mph}$$

for the velocity of the satellite.

2.4.4 Simple Harmonic Motion

This is the motion which is executed by all simple vibrating systems: a pendulum, the prongs of a tuning fork, a violin or piano string, the air column of an organ pipe, the water on the surface of a lake which is carrying a surface wave, the electric current in the antenna of a radio or television receiver which is responding to a broadcast wave, the effective electric charge and current in an atom as it radiates an electromagnetic wave, and many others. In fact, the motion we are about to describe is one of the most basic and universal of all the motions to be found in nature. If in fact a given motion is not precisely of this character, it may often be regarded as a superposition of several motions which are of this type. Consequently, it is of some importance that we define the type of motion we now have in mind rather precisely.

Our procedure will be to give a definition of simple harmonic motion (SHM), together with some of the immediate consequences of the definition, and then to show as we proceed that various physical situations which we encounter correspond exactly to the definition and therefore possess the properties of SHM. As in each of the previous cases the definition is given by making a definite statement about the acceleration. In this case the statement is the following: SHM is that motion for which the acceleration is proportional to the displacement and opposite in sign. This definition probably needs some amplification in order to be understood. Let us first simplify the situation by restricting the motion we are talking about to motion along a straight line. Then let us choose a reference point on this line from which to measure displacements. This should be a fixed point, and it can be taken as the origin of a coordinate system which in this case consists of a single axis, say the x axis. Finally, displacements in one direction, say to the right, can be taken as positive, and displacements in the opposite direction as negative. Now we are ready to understand the definition. If the system in its motion finds itself on the right at a certain distance from the origin, then according to the definition it will at that instant be experiencing an acceleration to the left, the magnitude of which is proportional to the distance from the origin. Similarly, if the displacement happens to be the left it will instantaneously be experiencing an acceleration to the right. If we put the definition in the form of an equation we will write

$$a = - Kx,$$

where x is the instantaneous displacement, a the instantaneous acceleration, and K the proportionality constant.

Some of the properties of SHM are immediately evident. Since the acceleration is opposite in sign to the displacement, the velocity is a maximum as the system passes through $x = 0$ in either direction and then decreases as x increases in positive or negative directions. Consider the situation as the system passes through $x = 0$ in the positive direction, i.e., with velocity to the right. As it moves further to the right the acceleration increases, and in the negative direction. This means that the positive velocity is decreasing at a rate that increases as the displacement increases. Eventually, the velocity must decrease to zero, and when this happens the system will have reached its maximum displacement on the right. We call this maximum displacement the amplitude of motion. But the acceleration is at its maximum negative value. Consequently, the velocity must continue to decrease which means that it is passing through zero toward negative values which then continue to increase. Negative velocity means motion to the left and consequently the system moves toward the origin and eventually passes through the origin with maximum velocity to the left. What happens after this is a repeat on the left side of what has just been described as taking place on the right side, all signs being changed from positive to negative and negative to positive.

It is clear that the motion is strictly periodic. We denote the period by T. It is the time lapse between corresponding points in two successive cycles; e.g., the time lapse between two successive passages through the origin moving in the same direction. Other quantities of interest are the maximum displacement in positive and negative directions (the amplitude of motion), which we denote by x_0, the maximum positive and negative velocity which we denote by v_0, and the maximum positive and negative acceleration which we denote by a_0. Also, in addition to the period of motion we may speak of the frequency ν. The period and frequency are related by

$$\nu = 1/T.$$

Further analysis of SHM shows that x_0, v_0, a_0, and T are related by

$$v_0 = \frac{2\pi}{T} x_0$$

$$a_0 = \left(\frac{2\pi}{T}\right)^2 x_0,$$

and that the period is related to the constant K in the defining equation for acceleration in SHM by

$$K = \left(\frac{2\pi}{T}\right)^2.$$

These relations are derived and discussed in the various textbooks on physics. See, for example, Holton, Introduction to Physical Science, pp. 99-102 (see footnote 2). The derivations will also be described in class.

3 DYNAMICAL PRINCIPLES

Motion in one form or another is the normal state of all matter. On the macroscopic scale we are aware that very large bodies are in a continual state of motion: the planets in their orbits around the sun, the satellites in smaller orbits around the planets, and galaxies in rotation and translation through the vast expanse of the universe. As we pursue our study we become aware of an equal persistence of motion on the microscopic scale: translational, vibrational, and rotational motion of all atomic particles constituting matter, systematic motions of electrons within atoms, and, on a still smaller scale, the motion of subnuclear particles within the atomic nucleus. In addition we are aware of a large array of natural and humanly controlled motions in our everyday experience. Obviously, there are many kinds of motion, involving many different types of particle and material bodies, interacting with each other in a great variety of ways. One naturally asks what it is that is common to all motion, what is its essence, what are the causal effects, and what are the principles by which these effects may be described and predicted.

3.1 DEVELOPMENT OF THE CONCEPT OF INERTIA

The answer to these questions came very late in the history of civilization. As we will now learn, the causal relations were not easy to discover. The difficulty lay in the fact that no body in nature is really isolated, and the achievement of effective isolation by experimental design can be realized only after the problem is well understood. The behavior of each body in any given situation consequently depends on unknown influences arising from the presence of other bodies. In the absence of explicitly stated principles, observations on the kind of motion occurring under a given set of conditions did not lead to conclusions with regard to the nature of the influences, and without an understanding of the influences the same observations could not lead easily to a discovery of the principles.

An essential preliminary step which had to be taken, first in the history of man, and now in the learning process of the individual student, is the development of an adequate conception of a property of matter which we call inertia. The simplest statement of this property involves an abstraction. We assert that any body which is not under the influence of any other body in nature, namely, a body completely isolated from all external influences, will remain in whatever state of rest or of uniform motion in a straight line in which it may have happened to be placed initially. This principle is a clear negation of the Aristotelian view, and a negation as well of the view of the scholastics, the followers of Aristotle at the end of the Middle Ages, who up to the time of Galileo asserted that a force is required to maintain an unchanging motion. To the Greeks and to the scholastics, any notion such as uniform motion in a straight line without the assistance of an agent to maintain the motion was simply preposterous.

The principle of inertia, essentially as described above, was first stated explicitly by Galileo, but hints with respect to it began to appear in the thinking of several of the natural philosophers during the immediately preceding centuries. Thus, Leonardo da Vinci in his long search for an understanding of motion came very close to expressing the spirit of the concept on several occasions. Before Leonardo, a fourteenth century Franciscan friar, William Ockham, discussed motion as an independent phenomenon, and he comes close to giving expression to a principle of inertia when he says that a body has motion because something of an abstract nature which he calls impetus has been imparted to it. This idea was carried further by his pupil Jean Buridan at the University of Paris, who went so far as to say that the moving object is carried forward by a nonmaterial property which is quantitatively equal to the product of the weight of the object and some function of its velocity. Successive restatements of these views during the next 250 years did not lead to further sharpening of this concept. There was, in fact, a deterioration of thinking in regard to it until, finally, the focus of ideas was again sharpened, and in the thinking of Galileo, and then Descartes, the concept of inertia was put upon a firm basis. It was essentially in the form as stated by Galileo that it was passed on to Newton for the next great and culminating advance.

3.2 NEWTON'S LAWS OF MOTION

The principle of inertia as described by Galileo explicitly states that in the absence of external influences of any kind a body continues to remain indefinitely in whatever state of rest or of uniform motion in a straight line it may happen to be. It says nothing about the manner in which changes in the state of motion may be achieved, or how those changes when they occur are to be related quantitatively to the influences which must be present. These influences may now be described as forces acting on the individual bodies. In the absence of forces all bodies continue to move at uniform velocity in a straight line or to remain at rest if initially there were no motion. It was Newton's great contribution to state correctly the connection between the forces which act on a body and the changes in motion which are a consequence of those forces. In doing this Newton founded what we generally speak of as the science of dynamics. The basic principles of this science are completely contained in Newton's three laws of motion which will now be described.

3.2.1 The First Law

This law is a restatement of Galileo's principle of inertia, with explicit reference to the absence of forces as the essential condition for straight-line uniform motion. In words which are very close to those used by Newton in the first published version of the three laws in the Principia, 1687, the first law is:

> Every body persists in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed upon it.

For the understanding of this law, it is of little consequence that hardly any natural motion having the specified characteristics exists. All natural motions are in curved paths of one kind or another, or in straight lines at continuously changing velocities, and we must conclude that in general forces are acting. What these forces are cannot be determined from the first law alone. If we do observe a motion which is at constant velocity in a straight line, as, for example, and automobile moving at constant speed on a straight highway, we must conclude that the force is zero. This may seem at first to violate our expectation based on common sense or intuition, but if we stress the implication contained in the statement of the first law that it is only the net force to which reference is made, the difficulty is removed. In the case of the automobile, for example, the object is being propelled forward by forces which have their origin in the power plant, but other forces which are a consequence of air resistance and friction are acting as well. These forces must be in the nature of opposing forces, and the constant speed of the automobile must be a consequence of the equality in magnitude of the opposing forces. Thus, the net force is zero and the motion is at constant velocity.

3.2.2 The Second Law

This law provides the connection between the magnitude and direction of a force and the changing state of motion of a body on which the force is acting. Newton observed that if in the absence of a force the velocity is constant, then in the presence of a force the velocity is changing, i.e., there is an acceleration. The simplest connection is that the force and the acceleration are simply proportional. We now note that the acceleration is a vector. This is proved in the same manner as we proved in section 2.4.2 that the velocity is a vector, starting, however, with $\Delta v^2 = \Delta v_x^2 + \Delta v_y^2$ instead of with $\Delta s^2 = \Delta x^2 + \Delta y^2$, for the case of two-dimensional motion. Since the acceleration is a vector, and the acceleration and force are proportional, then we anticipate that the force is a vector. Consequently, we now assert that the force and acceleration are proportional and in the same direction. This statement is the essence of the second law.

Before we can give a more explicit statement, we observe that there are two kinds of difficulties. One difficulty is that we do not in fact know what is meant by a force, and we do not yet know how to evaluate its magnitude. The second difficulty is that we clearly expect the connection between the magnitude of the force and the resulting acceleration to depend on a property of the body.

We will consider the second of these two questions first. The property of the body which is of interest here is its inertia. But the first law provides no measure of the inertia of any body. It says only that as a consequence of inertia, large or small, the body continues to move at constant velocity if it is not acted upon by any force. But intuition as well as the crudest of experiments tells us that different bodies have different inertias from the point of view of the magnitude of the forces required to change their states of motion, i.e., to give them an acceleration. It is an everyday experience that "heavy" objects are more difficult to set in motion, or to deflect from a predetermined state of motion, than "light" objects. Consequently, their inertial properties differ and we now look for a quantitative measure of inertia. We will eventually find this quantitative measure in the second law itself. For the moment we must recognize that our successive steps are to some degree intuitive and provisional, that we will find ourselves dealing with pairs of concepts which can have no independent meaning, and that the whole procedure can be justified only after all of the argument has been given.

It is from this point of view that we now assert that all bodies have measurable inertia and that the measure of this inertia is the mass of the body. The mass has not been previously defined. Naturally, we expect "heavy" bodies to have more mass than "light" ones, but we do not in fact know, and we will not assume, that mass and weight are proportional. Intuitively, we regard mass as a strict measure of the matter content of a body. Thus, the mass of a body will not depend on where in the universe it happens to be. The weight of course will; it varies with position on the surface of the earth, and with altitude, e.g. the weight changes in going from sea level to a mountain top.

Since mass is defined in terms of the matter content, a quantitative definition of mass can be given by starting with an arbitrarily chosen amount which may then be used as a unit in terms of which all other masses will be measured. What we use as a standard is simply a matter of convenience. We could have used the mass contained in a proton, or in a hydrogen atom if we had wished. Actually, we have found it quite convenient to begin with a clearly defined macroscopic quantity of matter. At one time we used the matter contained in one cubic centimeter of water at a temperature of 4° centigrade as the standard. We called this the gram mass. Subsequently, we have set aside another standard. This is a cylinder of platinum alloy which is preserved in the Bureau Internationale des Poids et Mesures at Sèvres, France, and duplicated for convenience in various bureaus of standards throughout the world. The quantity of matter in this platinum cylinder is very nearly one thousand times that contained in the cubic centimeter of water at 4° C. It is called the kilogram mass. It is the presently accepted standard throughout the world. The gram mass is now defined as the one-thousandth part of this standard.

We are now ready to state Newton's second law. We will state it as follows:

> The acceleration of a body is proportional to the net force acting and inversely proportional to the mass. The acceleration is in the same direction as the net force.

These are not precisely the terms in which Newton states the law. This is partly a matter of terminology. Some differences in content will be described in the appropriate places. For further information at this time about Newton and about his methods, together with excerpts from the Principia, see Shamos, Great Experiments in Physics (see footnote 1), and Cajori, Newton's Principia, Motte's Translation Revised.[3]

We will wish also to express the second law algebraically, and in addition, to make some reference to the units in terms of which the various physical quantities which enter into the law are to be measured. If we let m stand for the mass of the body, a for the magnitude of the acceleration, and F for the magnitude of the net force, then the second law can be expressed in the form

$$a \sim \frac{F}{m}.$$

This relation may equally well be written as $F \sim ma$. More conveniently still, let us introduce a proportionality constant k and write

$$F = kma.$$

The choice of k is arbitrary. Clearly, it defines the unit in terms of which F will be measured, having already fixed the units for m and a. A unit for F has not yet been chosen. Consequently we find it most convenient to define a unit in which $k = 1$. If we do this, the second law becomes

$$F = ma.$$

With m in grams and a in cm/sec^2, the unit for F is g-cm/sec^2. We call this unit the dyne. Using the equation we observe that the dyne of force is that force which gives to a mass of one gram an acceleration of one centimeter per second per second. The set of units in which mass is in grams, distances in centimeters, and time in seconds in called the cgs (centimeter, gram, second) system. In the cgs system, forces are in dynes.

We will conclude this portion of the discussion with two further remarks. The first is that the statement of the second law which has been given is presumed to be valid for all types of forces. The force acting on a body in any given case may be gravitational, electrical, magnetic, nuclear, or simple mechanical (as the push or pull

[3]Florian Cajori, Newton's Principia, A. Revision of Motte's Translation (University of California Press, Berkeley, 1934).

of the hand or as a consequence of the operation of a mechanical device). No matter what the nature of the force, the second law as stated here is presumed to be the relation which provides the correct description of the resulting motion. The proper description of these forces in terms of their ultimate causes is a separate problem, but once they have been described and evaluated as forces of defined magnitudes and directions, Newton's second law is available as a basis for the description of the resultant motion.

The second remark is that we can now show that Newton's second law is available for the measurement of mass. In principle, this is accomplished by using the standard mass in order to calibrate a set of forces. This is done by measuring the acceleration of the standard mass when acted upon by the forces. If the standard mass is 1 g, the measured acceleration directly gives the value of the force in dynes. Now use the measured forces to accelerate unknown masses. According to the equation of motion, the masses in each case are equal to the forces applied divided by the measured accelerations. Consequently, any mass may be measured, and the proportionality of acceleration to force may be checked under a large variety of circumstances. Two additional methods of measuring masses, one based on Newton's second law and one independent of it, will be described later.

3.2.3 The Third Law

The simple forces in nature, and the ones of greatest interest to us, are those which act between pairs of bodies and along a straight line drawn from one to the other, e.g., the mutual attraction of the earth and sun for each other must be viewed as an attraction of the sun for the earth as well as the earth for the sun; a body resting on a table top pushes down against the table and the table pushes up against the body; an automobile accelerating on a highway pushes against the road and the road pushes against the automobile to give it the acceleration. Thus all forces are in effect double-ended. Newton's third law states this important property of forces in concise terms. It may be expressed as follows:

> In the mutual interaction of two bodies with each other, the force on the first body due to the second is equal and opposite to the force on the second due to the first.

As is generally the case with the laws of physics, the range of validity of Newton's laws can be determined only as a consequence of detailed study of the implications and by comparison of the predictions with observational results. Applications of the laws to simple physical situations and a description of some further implications will be given in the next few sections. Two remarks in regard to validity may be made at this time, however. The second law is obviously valid only for observers who are themselves not being accelerated. In the words of Newton, the observer must be in an inertial system. He defined an inertial system as one which is not accelerating with respect to the fixed stars. With this restriction the second law is found to have general validity for all motions which occur at low velocity. The modification of the second law which is required when the velocity is not small, i.e., not small compared with the velocity of light, will be considered later. The question of the range of validity of the third law is a more difficult one. We will make use of it only for objects which are at rest with respect to each other or which are in actual contact. Under these circumstances the third law is found to have complete validity and the conclusions which we draw will correspond to the facts of nature. One of these conclusions is the conservation of momentum to be described below. We will then find that momentum conservation can be derived on the basis of a general principle of physics which we will state: the relativity principle. We will also observe that the validity of momentum conservation is easily checked experimentally, and we will anticipate the fact that no violations of momentum conservation have ever been observed. In effect, then, Newton's third law, which is known to have limited validity, will be replaced by the momentum conservation principle as the required third law of dynamics.

3.3 MOTION IN THE GRAVITATIONAL FIELD OF THE EARTH AT THE EARTH'S SURFACE

The characteristics of all motion at or near the surface of the earth in which there are no forces acting except the force of gravity are well-known from Galileo's work. As we have seen, he characterized this motion as motion at constant acceleration g vertically downward and at zero acceleration horizontally. The only thing we learn from an application of Newton's second law to this situation is the magnitude of the force of gravity on any given object.

Consider the free fall of an object of mass m. It falls with an acceleration g, the numerical value of which can be determined in any given locality. Then, since a = g, Newton's second law gives for the force of gravity F_g the result

$$F_g = mg.$$

Since $g = 980\ \mathrm{cm/sec^2}$ approximately, the force of gravity on a one gram mass is 980 dynes. Note that this is an experimental determination of the force of gravity on mass, and we do not yet have any theoretical basis for describing this force.

The following problems involving applications of Newton's second law to physical situations in which the only force acting is the force F_g will be of interest. In addition, the last of these problems, the problem of the simple pendulum, provides an excellent method for the precise measurement of g.

3.3.1 Object Sliding On An Inclined Plane

By the use of inclined planes, Galileo was able to reduce the effective acceleration of gravity in a manner which we will now demonstrate. Let a mass m rest on a smooth flat surface inclined at an angle θ to the horizontal. Assume that the object can slide without friction on this surface. This condition may be achieved approximately in practice by using a piece of smooth dry ice as the mass m and a sheet of plate glass for the plane. A very much improved technique, involving materials which are now commercially available, is to use a linear air trough in which the sliding object is continuously supported as it moves on a film of air. For a description of this device, see H. V. Neher and R. B. Leighton, Am. J. Phys. 31, 255 (1963).

Referring now to Fig. 3.1, we identify the force of gravity acting on m, and we resolve this force as indicated into two components, one parallel to the plane, which we denote as F_1, and the other perpendicular to the plane, which we denote as F_2. From the geometry one observes that $F_1 = F_2 \sin\theta$ and $F_2 = F_g \cos\theta$ where $F_g = mg$.

The only other force acting on m is the reaction of the plane against m. The force of reaction, like the force of gravity, may be resolved into

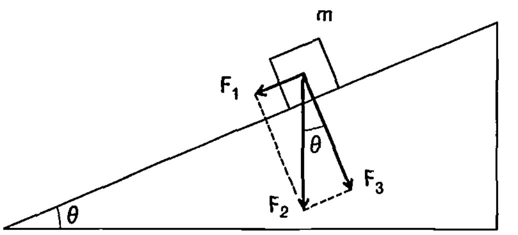

Fig. 3.1

parallel and perpendicular components. But the parallel component is equal to zero because the motion is specified as frictionless. The perpendicular component must be equal and opposite to F_2 since the body has no acceleration perpendicular to the plane. Only one force remains; namely, F_1. This produces acceleration down the plane. Setting $F_1 = ma$ one obtains

$$a = g \sin \theta .$$

We observe that the acceleration can be made arbitrarily small by making θ small. Also the acceleration is a constant, and the kinematical relations valid for constant acceleration may be used to describe the position and velocity as a function of time starting with any desired initial conditions. The student should make up several problems, some with numerical data and describe the motion in each case.

Perhaps it should be mentioned that Galileo used rolling objects instead of sliding objects. It was in this way that he was able to reduce the effects of friction to a negligible value. Further analysis of this problem would show that for rolling objects as for sliding objects the conditions of constant acceleration obtain, and Galileo was correct in his conclusion that he had demonstrated the existence of constant acceleration in this motion by observing that the ratio of distances traveled in successive units of time were the ratios of the odd integers. However, he could not have predicted the actual accelerations since he did not have Newton's second law available. Also, we are not yet ready to predict the acceleration for rolling bodies. In order to do so, we must make a further study of Newton's equation and its application to bodies which rotate and to bodies which rotate and translate simultaneously. Such topics will be treated in another chapter.

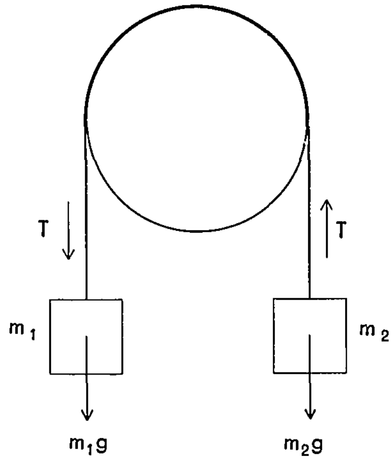

Fig. 3.2

3.3.2 Atwood's Experiment

A physical arrangement of some interest is the one shown in the diagram (Fig. 3.2), where it is assumed that the wheel over which the string is hung rotates without friction and that its inertial properties are negligible.

If the two masses are unequal, say $m_2 > m_1$, there will be an acceleration because the net force on the system is not zero. The acceleration is easily determined by noting that the net force is $F = (m_2 - m_1)g$ and that the total mass to be accelerated is $(m_1 + m_2)$. We assume that the string is inextensible and that as a consequence the two masses have precisely the same acceleration, one up and one down. Then using Newton's second law,

$$(m_2 - m_1)g = (m_2 + m_1)a$$

or

$$a = \frac{m_2 - m_1}{m_2 + m_1} g .$$

Again the motion is at constant acceleration and the kinematical relations derived for this case are available for use. Instead of considering questions of this type, however, let us note that in this case we have another question that we could ask; namely, what is the tension in the string. The answer to this question is also obtained by an application of

Newton's second law. Consider for example the mass m_2. The two forces acting on it are $m_2 g$ downward and the tension T in the string upward. Thus the net force on m_2 downward is $m_2 g - T$. Substituting into the equation of motion, F = ma, we have

$$m_2 g - T = m_2 a.$$

Since a is already determined, we can introduce it into this equation and solve for T, thus obtaining

$$T = m_2 (g - a) = \frac{2m_1 m_2}{m_1 + m_2} g .$$

3.3.3 The Simple Pendulum

If a mass m which is essentially a point particle is attached to a string and is allowed to swing freely under the action of gravity when the other end of the string is attached to a fixed point, we have an arrangement which we call the simple pendulum.

Denote the length of the string by ℓ, and consider the system at an instant when the motion is to the right and the string makes an angle θ with the vertical. This is illustrated in Fig. 3.3: where we have also indicated that the actual distance between m and the vertical line is x, and that the displacement along the arc is S. We again resolve the force of gravity mg into two components, one of which is parallel to the direction of the string and has the effect only of producing tension in the string, and the other is $F_1 = mg \sin\theta$ which is directed along the arc and is responsible for the acceleration. Using the equation of motion $F_1 = ma$ where a is directed along the arc, i.e., in the same direction as F_1, we find

$$mg \sin\theta = ma$$

or

$$a = g \sin\theta.$$

But from the figure $\sin\theta = x/\ell$. Also, for small amplitudes of motion to

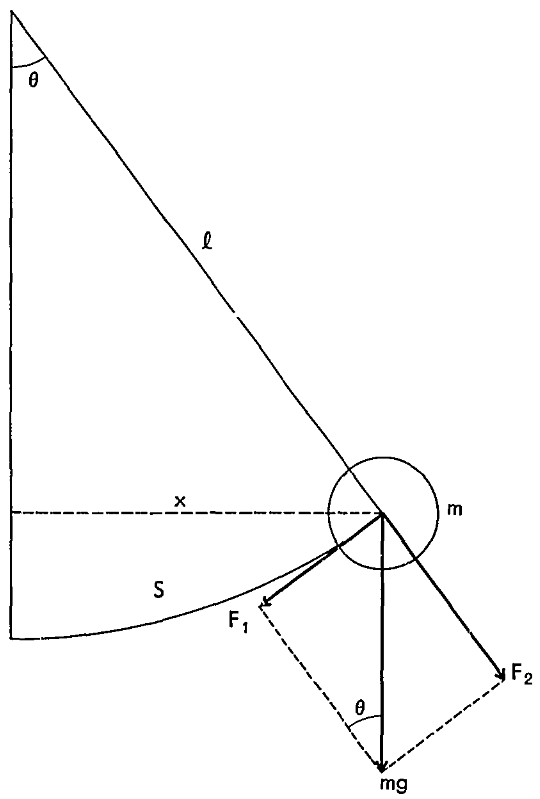

Fig. 3.3

which we now restrict our discussion $x \cong S$. Then

$$a = \frac{gS}{\ell} .$$

But when S is to the right (positive), a is to the left (negative). Therefore, the final expression for the acceleration in small amplitude motion is

$$a = -\frac{g}{\ell} S = -kS,$$

where $k = g/\ell$.

Since a is proportional to the displacement and opposite in sign, the motion of the simple pendulum is simple harmonic. But for SHM we have learned that the period and the constant K are connected by the relation

$$K = \left(\frac{2\pi}{T}\right)^2 .$$

Solving for T and substituting $K = \frac{g}{\ell}$ we find

$$T = 2\pi\sqrt{\frac{\ell}{g}}.$$

Since ℓ and T can be measured with precision, g is easily determined.

3.4 LINEAR MOMENTUM AND MOMENTUM CONSERVATION

The product of mass and velocity turns out to be an especially important physical quantity. It is called the momentum and we denote it by the letter p. Thus

$$p = mv.$$

Since v is a vector, p is a vector. Newton recognized the special physical significance of the momentum. He called it the quantity of motion.

Since the mass is constant (at least in the context of the present discussion), we can use the definition of acceleration to write Newton's equation of motion in a different form; thus

$$F = ma = \frac{m\Delta v}{\Delta t} = \frac{\Delta(mv)}{\Delta t} = \frac{\Delta p}{\Delta t}.$$

In words, force is the time rate of change of momentum. This is in fact the way in which Newton originally stated his second law. Our present interest in writing the equation of motion in this form is that we can write an expression for the change of momentum as a product of the force and the time interval over which it acts. Thus

$$\Delta p = F\Delta t.$$

This equation certainly holds for constant forces over time intervals of arbitrary length. It also holds for variable forces over short time intervals during which the force doesn't change very much during the interval and an intermediate value may be used. If the force is variable and the time interval which we wish to consider is not a short one we can divide the long time interval into a large number of short intervals in each of which we regard the force as constant. Calculating the products $F\Delta t$ for each time interval and adding them we obtain a result which is in effect the product of the average force over the whole time interval and the time during which it acts. We call this quantity the impulse and denote it by the letter I. Thus Newton's equation of motion becomes

$$\Delta p = I,$$

or, more explicitly, if the particle of mass m has the velocity v_0 at the beginning of the time interval, its value v at the end is given by the relation

$$m(v - v_0) = I,$$

where the impulse I can be evaluated by adding up all of the contributions to I which occur during the short intervals into which the whole interval is divided.

The importance of this result lies in the fact that one can discuss the total change of momentum without discussing the details of the way in which the force varies as a function of time. An important case is that of collisions between two particles. Let the masses of the two particles be denoted by the letters m and M. Let the corresponding velocities before the collision be denoted by v_0 and V_0, and after the collision by v and V. For simplicity, the motion of both particles will be confined to the same straight line. Velocities which are positive will be to the right and velocities which are negative will be to the left. Now consider what happens during a collision. A variable force

F(t) is exerted by one particle on the other and as a consequence of the collision a total impulse I will have been transmitted. But by Newton's third law an equal and opposite force will be acting between the two particles in the opposite direction; i.e., on the other particle, and it acts for the same time. Thus, the impulse transmitted to the second of the two particles is −I. Then, for the two particles of mass m and M we have

$$m(v - v_0) = I \text{ and } M(V - V_0) = -I.$$

Combining these two equations we obtain

$$m(v - v_0) = -M(V - V_0)$$

or

$$mv_0 + MV_0 = mv + MV.$$

This result has a simple interpretation. The left-hand side of the equation is the sum of the two momenta of the particles before collision; the right hand side is the sum of the momenta after collision. The two sums are equal. Consequently, the momentum of the system has not changed. This is the simplest case of the general theorem that the total momentum of an isolated system remains constant independent of all interactions within the system. The extension of the proof to the case of motion in three dimensions and to systems containing more than two particles presents no difficulties.

We can use the principle of momentum conservation to measure mass. Let m by a standard mass and M an unknown mass. Allow a collision to take place with m moving initially at velocity v_0, M initially at rest ($V_0 = 0$), and measure v and V after collision. Then from the momentum conservation equation

$$M = m \frac{v_0 - v}{V},$$

where all quantities on the right side are known. Other examples of momentum conservation will be described in class.

3.5 GALILEAN RELATIVITY

In the preceding section the principle of conservation of momentum was proved using a combination of arguments based on Newton's second and third laws of motion. Since momentum conservation is believed to have universal validity whereas the third law will be shown to have a limited validity, it is of considerable interest, and a source of satisfaction, that we are able to find a proof of it which is based on a principle of physics that has general acceptance and is unrestricted in its applicability. This is the relativity principle which states that the laws of physics look the same to all observers who are moving with constant velocity with respect to each other.

In order to relate observations made by one observer with those made by another, one has to specify the manner in which the two systems of observations are to be related. The intuitively correct way of making a comparison between two observers who are moving at constant velocity with respect to each is simply to add or subtract this constant velocity from the velocities measured by one of the two observers. Thus, if observer A is moving with respect to B at constant velocity u and A says that something which both are observing has velocity v, then B will say that it has velocity v + u. When sense data are related between two systems in this way, we say that they are being related by a Galilean transformation, and the relativity principle which is being used subject to this kind of a transformation is called Galilean relativity. We are mentioning these terms at this time because we will find later that another relativity principle and transformation must be considered if we wish to deal with motions at high velocity.

We note that the form of Newton's second law is unchanged in a Galilean transformation, and consequently this is one law of physics that can be immediately verified to be the same for two observers moving at constant velocity with respect to each other. The proof is simple. Each of two observers A and B uses the second law in the form

$$F = m \frac{\Delta v}{\Delta t},$$

where v is the change of velocity in a time Δt of a mass particle that both A and B have under observation. Since the actual instantaneous velocities which each observes differ only by the constant relative velocity u which they have with respect to each other, they will agree on all change of velocity Δv. Consequently, each correctly predicts the motion in his own system using the same law of motion. Simple examples in everyday experience illustrate this result. A child may bounce a rubber ball on the floor of an airplane in flight in the same way that it bounces the ball on the floor of its home. An object dropped from mast height of a moving ship in quiet water hits the deck at the same point as when the ship is stationary. Without knowledge of Newton's laws of motion, Galileo used the relativity principle in answer to critics who asserted that if the earth was really in motion as specified by the Copernican theory, an object dropped from the top of the tower of Pisa should not land at its base. He demonstrated the actual motion and explained why it occurs in this way.

We now turn to the proof of momentum conservation using only the relativity principle. This proof can be based on very simple experiments. We consider the motion of two or more mass particles which move without friction along a single straight line. The near absence of friction is achieved by using the air trough described in Section 3.3.1 as the track along which the objects move. In all cases the motion is initiated by having two mass particles in contact and permitting a small explosion to take place between them. Then the objects move away from each other at constant velocities (since there is no friction), which are measured, and then after a certain time they are reflected back from the ends of the track by suitable reflectors which are placed there. They eventually come into contact again and it can be arranged that on making contact they stick together. The nature of the subsequent motion depends on the circumstances. The desired data consists in the measurement of the velocities at all stages of the motion. The relativity principle is used to obtain predictions about the motion in one frame of reference when it is known in another, and thus to obtain results in certain nonsymmetrical situations from results which are easily obtained in symmetrical situations. In the course of the analysis one finds that a definition of mass based on a a comparison of measured velocities can be made which is independent of the kinds of matter in a body and does not depend on velocity (for these cases of small velocity). One also finds that as a consequence of the analysis one has proved the principle of conservation of momentum.

A complete discussion and demonstration of associated experiments will be given in class. Students should consult Feyman, Leighton, and Sands, Lectures in Physics, Vol. I, pp. 10-3 to 10-7,[4] in advance of this discussion.

3.6 ANGULAR MOMENTUM, TORQUE, AND ANGULAR MOMENTUM CONSERVATION

It is frequently of interest to describe the motion of a point parti-

[4] Richard P. Feynman, Robert B. Leighton, and Matthew Sands, Lectures on Physics (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1963).

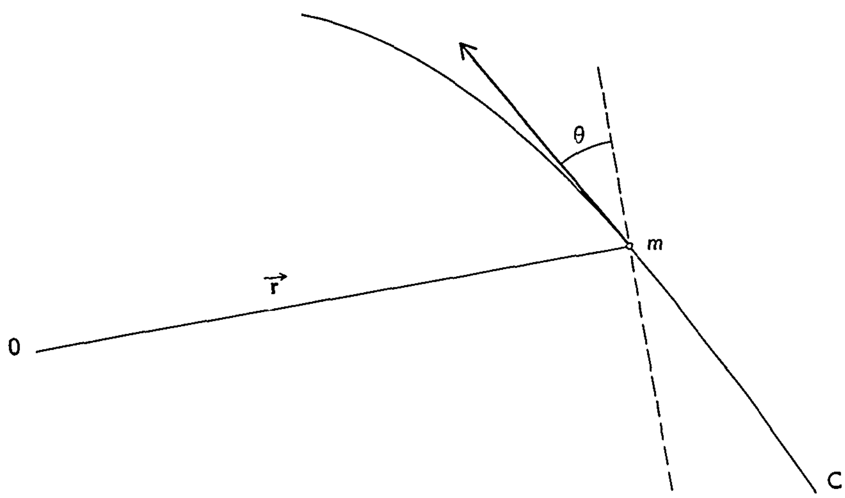

Fig. 3.4

cle with respect to a fixed point without restriction as to the kind of motion the particle may have. At a particular instant of time, let the particle be at the vector distance $\vec{r}$ from the fixed point 0 moving along a path C at the instantaneous velocity $\vec{v}$.

The velocity vector $\vec{v}$ makes the angle θ with respect to a line perpendicular to the vector $\vec{r}$ connecting the point 0 to the instantaneous position of the particle (Fig. 3.4).

We now define a quantity which we call the angular momentum and denote i, by the sumbol A. The definition is

$$A = mvr \cos \theta = mrv_{\perp} = mvr_{\perp},$$

where $v_{\perp} = v \cos \theta$ is the component of $\vec{v}$ perpendicular to $\vec{r}$, and $r_{\perp} = r \cos \theta$ is the component of $\vec{r}$ perpendicular to $\vec{v}$. The angular momentum is regarded as positive if the rotation about 0 is counterclockwise as seen from above the page. These relations provide useful alternative definitions of angular momentum. It is convenient to have still another definition. The instantaneous motion is in a plane which contains the vectors $\vec{v}$ and $\vec{r}$. Consider an xy coordinate system in that plane. If the coordinate system is oriented so that the positive x axis points along $\vec{r}$, the above definitions of A give

$$A = mv_y x.$$

If, on the other hand, the coordinate system is oriented so that the positive y axis points along $\vec{r}$, the definitions yield

$$A = -mv_x y.$$

Then in general for an arbitrary orientation of the coordinate system

$$A = m(v_y x - v_x y).$$

We are now going to show that the angular momentum changes with time in a manner which is determined uniquely by the forces and the position of m with respect to 0. Consider first a small change in A which is regarded as occurring in a small time interval Δt. Since each term of A is a product of two factors, each of which changes with time, we obtain for the change of A

$$\Delta A = m(v_y \Delta x + x\Delta v_y - v_x \Delta y - y\Delta v_x).$$

Dividing by Δt, we obtain the rate of change of A. Thus

$$\frac{\Delta A}{\Delta t} = m\left(v_y \frac{\Delta x}{\Delta t} + x \frac{\Delta v_y}{\Delta t} - v_x \frac{\Delta y}{\Delta t} - y \frac{\Delta v_x}{\Delta t}\right)$$

$$= m(xa_y - ya_x),$$

where we have used

$$v_x = \frac{\Delta x}{\Delta t},\ v_y = \frac{\Delta y}{\Delta t},\ a_x = \frac{\Delta v_x}{\Delta t},\ a_y = \frac{\Delta v_y}{\Delta t}.$$

Now consider the force acting on m. At a given instant the force $\vec{F}$ has the components F_x and F_y in the chosen xy coordinate system. By Newton's second law of motion

$$F_x = ma_x \text{ and } F_y = ma_y.$$

Let α be the angle between the direction of $\vec{F}$ and the perpendicular to $\vec{r}$ (just as θ is the angle between the direction of $\vec{v}$ and the perpendicular to $\vec{r}$). Then define the torque L by the relations

$$L = Fr \cos \alpha = F_{\perp}r = Fr_{\perp},$$

which is to be regarded as positive if it tends to produce rotation in the counterclockwise direction. By the same argument as before this corresponds to

$$L = F_y x,$$

if the positive x axis is oriented along $\vec{r}$, and to

$$L = -F_x y,$$

if the positive y axis is oriented along $\vec{r}$. Then in general for any orientation of the coordinate system

$$L = F_y x - F_x y.$$

Using the equations of motion

$$L = m(xa_y - ya_x).$$

Comparing with the expressions for $\Delta A/\Delta t$, we obtain the simple general result

$$L = \frac{\Delta A}{\Delta t}.$$

This is the form taken by Newton's second law when expressed in terms of torques and angular momenta rather than in terms of forces and linear momenta. It has the immediate consequence that if the torque on a particle about any point is zero, the angular momentum is a constant. In this form the statement is analogous to Newton's first law for linear motion. However, it can be immediately generalized. If there are several particles, or more generally, any system of particles interacting internally in any way whatever, but no net torque on the system as a whole, the total angular momentum of the system remains constant in time. This general result is of great importance in physics. It is the theorem of conservation of angular momentum. In this form it is the rotational analog of the previously proven theorem of conservation of linear momentum. Examples illustrating the principle of conservation of angular momentum will be described in class.

3.7 ENERGY AND ENERGY CONSERVATION

Energy is one of the most fundamental of all the concepts of physics, and in a certain sense it is also one of the most abstract. Partly for this reason a simple unambiguous definition is not possible at the beginning of our study. The most that we can do is to define certain forms of energy, and to describe certain relations which suggest why the concept is of some importance. As we proceed, we will find that there are other forms of energy which are to be added to the list of forms already known. Also we will find that an outstanding characteristic of physical processes is that energy is being continuously transformed from one form to another, and most important of all, that this transformation occurs without any loss or gain in the total energy which is present in a given system, or, if there is a loss or gain, the difference can be accounted for in terms of a gain or loss in outside systems. This leads to another conservation law, a law as fundamental to the understanding of physical processes as the law of conservation of momentum, but of an entirely different nature. Whereas we have been able to give a simple definition of momentum and the associated momentum conservation law, the definition of energy and its conservation will have to be developed as we proceed. We are here making a beginning of this study.

Consider a constant force F acting on a mass particle m. By Newton's second law of motion, the acceleration is a constant and is determined by the relation

$$F = ma.$$

For simplicity let the particle be restricted in its motion to a straight line and let the force act along this line. Multiply both sides of Newton's equation by S, the distance through which we choose to observe motion of the particle while acted upon by the force F. This gives

$$FS = maS.$$

Since the motion is at constant acceleration, we can replace the product (aS), on the right hand side of the equation by using the relation

$$v^2 - v_0^2 = 2aS.$$

This gives

$$FS = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mv_0^2.$$

We note carefully the implications of this equation. The only factors appearing on the left-hand side are the applied force and the distance through which the force has been allowed to act. The quantities appearing on the right-hand side refer only to the state of motion of the particle; i.e., in addition to the mass they depend only on the velocity of the particle. Furthermore, the right-hand side is a difference of two terms of the same kind, one an initial value, and the other a final value. This suggests a kind of conservation law, and we now interpret the result in this way. We will call FS the work done by the applied force F and denote it by W. We will call $\frac{1}{2}mv^2$ the kinetic energy of motion of the particle and denote it by K. Then the relation we have derived can be written,

$$W = K - K_0,$$

which says that the change of kinetic energy of the particle is equal to the work done by the applied force. In other words we have made a definition of a new dynamical property, and at the same time a definition of effort by an outside agency, in such a way that the effort expended by the agency turns out to be equal to a gain in the magnitude of a certain property possessed by the body.

We will find that the definition of work which we have made here can always be used. If the applied force is not in the direction of motion, we generalize the definition so that it reads product of force and the component of displacement along the direction of the force or product of displacement and the component of force along the direction of the displacement. The student will show easily that these two forms of the definition are exactly equivalent. The definition may be written symbolically in the form

$$W = FS \cos \theta,$$

where θ is the angle between the directions of the force and displacement. A further generalization is to the case that the force is not constant with change in displacement. Then the latter must be regarded as divided into small intervals within each of which the force may be regarded as constant. By evaluating the work done in each of the small intervals and then taking a sum over all intervals, one obtains the work done.

As we have seen, a body possesses energy by virtue of its motion. This energy is called kinetic energy. Another form of energy which the body may possess is by virtue of its position in space and this energy is called potential energy. Consider a particle in a force field, e.g., a mass particle m in the gravitational field of the earth, and consider displacements such that the particle is never far from the surface of the earth. Then the force $F = mg$ is approximately the same for all positions and we say that the particle is in a constant force field. In other examples the force will vary with position in space, but whether it varies from point to point or not is of no consequence for our present purpose. The main point is that there is a force and that this force is a function of position in space (including also the case where it is a constant). All such situations are characterized by the statement that the particle is in a force field.

We now observe that work must be done by some unspecified agent if the particle is to be moved from one point in a force field to another. We call the work done the change of potential energy of the particle in the force field between the two points in question. We denote the potential energy by U. If the agent moves the particle along the line of force in the force field, but in the opposite direction, a maximum work is done per unit distance of separation of the two points, and it is positive. We say that there is an increase of potential energy. If the agent moves the particle perpendicular to the direction of the lines of force in the force field, no work is done, and there is no change in potential energy. In calculating the change of potential energy between any two points in the force field, the force applied by the unspecified agent must at all times be just equal and opposite to the force in the force field plus a very small increment ΔF to cause eventual displacement from one point to the other. If $\Delta F = 0$, there is no displacement, e.g., a mass m supported on a table top is at rest because the force mg downward in the force field is precisely balanced by the force mg upward exerted by the outside agent, the table top. If ΔF is not zero, there is motion with acceleration. In order to calculate changes of potential energy between two points, we regard ΔF as being so small that the acceleration, and therefore the gain in kinetic energy, is negligible, and in the limit of $\Delta F = 0$, is actually zero.

We speak of difference of potential energy between two points in space rather than potential energy at a point because there is no unique reference point from which to measure potential energy. However, we frequently define one. For example, if we are interested in the motion of a mass m with respect to the floor in our laboratory, we would define the floor as the position of zero potential energy. Then all points above the floor are points of positive potential energy. For example, the potential energy of the mass m at a height h above the floor is

$$U = mgh,$$

because the work done in moving the object from the floor to the point in question is the product of the force mg which one must apply and the height h through which the object is moved. The choice of a zero of potential energy is always a matter of convenience and it will be made in quite different ways in the various situations which we will encounter.

We will now summarize our ideas on energy insofar as we have gone. A mass particle possesses kinetic energy because of its velocity of motion and it possesses potential energy depending on where it is in a force field. If an unspecified agent does work on the mass particle, this work may be used to change the kinetic energy, or to change the potential energy, or to do both simultaneously. In general

$$W = \Delta K + \Delta U.$$

We also note that if there relation is correct for a mass particle which is being acted upon by an agent, it must also be correct in the limit that the agent does no work, i.e., $w = 0$. Then for the isolated mass particle in a force field

$$\Delta K + \Delta U = 0.$$

This is the simplest form of the energy conservation law. It says that a mass particle always moves in such a way that the sum of its kinetic and potential energies is a constant.

These concepts will be developed at greater length in class, and applications to simple situations will assist in the clarification of the ideas involved.

4 FORCE

4.1 INTRODUCTION

In our consideration of the laws of motion we have dealt so far mainly with the inertial properties of matter, and with certain consequences of the second law, such as the conservation of energy, the validity of which may be assumed provisionally to be independent of the nature of the forces. Most important has been the result that a quantitative comparison of inertial masses, independently of the nature of the matter involved, can be given. Of considerable interest also was the fact that the comparison of masses could be based separately on distinct and seemingly independent principles of physics: on the laws of motion on the one hand, or, alternatively, on a general principle that all of the laws of nature are essentially the same for observers in different inertial systems, i.e., no one inertial system is to be regarded as preferred over any other. Thus, the principle of inertia as embodied in the statement of the first law, and the concept of inertial mass as used in the second law would appear to be on a rather sound foundation.

But we have not yet had any real test of the second law. Most of the applications which have been made so far have dealt with motions of objects under the influence of a gravitational force near the earth's surface in which we already had full knowledge of the acceleration, at least for the case of free fall. As we have seen, the only role of the second law in such cases was merely to extend the terminology, to permit a discussion in terms of forces as well as in terms of accelerations. Thus a constant acceleration was discussed as a constant force; nothing new about the motion was learned. One exception was the application of the second law to the motion of a particle at uniform speed in a circle. In this case we used the second law to predict the magnitude of the force required to keep the object moving in a circle, and we compared this calculated force experimentally with the force of gravity on the same object. In this way we showed that the measurement of a force, using Newton's second law, gave results in two entirely different physical situations which were consistent with each other. A few other examples demonstrated also the nontrivial nature of the concept of force, but, on the whole, we learned nothing new about the nature of force.

In order to go further we must say something specific about forces. If we are to predict the kind of motion which occurs in any given physical situation, we must know not only that force and acceleration are related to each other in a certain way, but we must have also detailed knowledge of the forces which are acting. This information about forces is obviously independent of and supplementary to our knowledge of how forces act to change the state of motion. In effect the information which we require about forces constitutes the basis for the statement of separate laws of physics, and these laws must be ascertained, tested, and provisionally verified in much the same way as all other laws of physics.

Much of the history of physics since the beginning of the scientific revolution in the seventeenth century has been concerned with the determination of force laws. It began when Newton, contemplating the motion of the planets and searching for a test of his second law of motion, asked himself about the forces which must exist between the planets and the sun. As we will see, he came up with a law, the law of universal gravitation, describing one of the basic interactions in nature.

The discovery of other types of interaction occurred in the eighteenth, nineteenth, and twentieth centuries. At the present time we know of four basic interactions. According to our present knowledge, all of the forces of nature, nuclear and subnuclear, atomic and molecular, the forces responsible for planetary and galactic motion, and all of the forces of everyday experience, are made up of one or more of the four basic interactions. Thus, it is in terms of basic interactions that we will eventually describe all of the natural motions, e.g., electrons in atoms, neutrons and protons in the atomic nucleus, atoms and molecules in macroscopic matter, and the planetary and galactic motions of the large scale universe. It is also in terms of these basic interactions that we explain how matter is held together, and why it is not infinitely compressible. Also we will be able to convince ourselves that all of the forces of everyday experience, e.g., the forces between the various components of a mechanical device, the forces involved in the contraction of a muscle, and the forces between the hand and a rigid object which it grasps and seeks to move, these forces and all other forces which might be mentioned, can be described in terms of the basic interactions. Clearly the basic interactions are of fundamental importance. They will be mentioned and briefly characterized in the next section, and then described in more detail in succeeding chapters. Because of its historical importance and its relevance in connection with the first real test of the validity of the second law of motion, special consideration will be given to the law of universal gravitation and its applications to a wide range of planetary, terrestrial, and interplanetary motions. This will be the subject of the next chapter.

We also have need in our study of physics for force laws which are determined empirically. It is in fact through the continuous interplay of empirically determined force laws and a similtaneous search for an understanding in terms of general principles that the nature of the basic interactions was actually discovered. Thus, our knowledge of the nature of the electromagnetic interaction was obtained from observations made on large collections of static electric charge and on the interaction forces between metallic wires carrying electric currents rather than on the forces between the members of a single pair of charged particles moving in a specified way with respect to each other. But it is generally not a simple matter to deduce the basic interactions from the forces as observed in a specified macroscopic situation, or, conversely, to find the form of the large-scale interaction by performing a summation over all of the basic interactions which are present. For this reason the student of physics needs to study both aspects of the problem simultaneously. He studies the basic interactions in order to gain insight into the nature of physical processes by describing those which are elementary in considerable detail, and he studies the behavior of various systems in the presence of empirically determined forces in order to achieve an understanding of the different kinds of motion which occur in complex physical situations. It is important, however, to realize that in principle one set of forces can be determined from the other, and in more advanced courses the student learns how this is done. A description of several of the more important and interesting of the empirically determined forces is given in the section following the description of the basic interactions.

4.2 THE FOUR BASIC INTERACTIONS

In the following section we briefly characterize each of the following:

(a) The gravitational interaction.
(b) The electromagnetic interaction.
(c) The nuclear interaction; also called the strong interaction.
(d) The weak interaction.

All the forces of nature fall into one or the other of those four categories. The following description is intended to provide some initial orientation about them. Each must then be studied in considerable detail in relation to the physical phenomena of which each is the principal part.

4.2.1 The Gravitational Interaction

This is the simplest of all the interactions and it was the first to be discovered. All matter in the universe experiences a force of attraction for all other matter. It is assumed that the magnitude of the force is proportional to the product of the masses of the interacting bodies and inversely proportional to the square of the distance between them. The direction of the force lies along the line connecting the two bodies, and according to Newton's third law of motion, each of the two bodies, when at rest with respect to each other, experiences the same attracting force toward the other. In symbols the magnitude of the force is given by

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

where m_1 and m_2 are the two masses and r is the separation distance. An important part of the assumption is that the constant G is a universal constant; i.e., the force of attraction between two masses at a given distance of separation is independent of where the masses are in the universe. It is on the basis of this assumption that we denote the law as the law of universal gravitation. However, the asumption of proportionality to mass does not necessarily imply that the mass with which we are now dealing is the same as the inertial mass appearing in the second law of motion. Strictly speaking we should call the mass to which the force of gravitational attraction is proportional the gravitational mass. The connection between gravitational and inertial mass must then be determined by experiment. In recent years this connection has been studied under conditions such that differences between gravitational and inertial mass of as little as 1 part in 10^{10} would be detected. No difference has yet been observed. Consequently we assume that inertial mass and gravitational mass are proportional, and we will choose the units such that they are equal.

We do not know precisely when and how the idea of a law of universal gravitation first came to Newton. The law was first published in 1687 in the Principia, a great work containing all of Newton's original contributions to science, mathematics, and philosophy. Newton himself says that the law of universal gravitation came to him when he asked himself whether the falling of an apple to the earth's surface could be caused by the same kind of force that causes the moon to fall steadily toward the earth as it moves forward with a constant linear speed in its circular orbit. He then calculated the moon's acceleration from the known period and radius of its orbital motion, and, in a manner similar to that which we described in the section on kinematics, he noted that the acceleration of the moon was smaller than that of the apple (or any other falling body at the earth's surface), by a factor which was approximately equal to the ratio of the squares of the distances to the center of the earth. In the words of Newton he "found [the two forces, on the moon and on the apple] to answer pretty nearly." This was in 1665 and 1666, twenty-one years before the publication of the Principia. It suggested to him the inverse square law of force. The proportionality to mass, which Newton assumed from the beginning to be the inertial mass, then follows from the fact that accelerations at the earth's surface are the same for all bodies, and consequently the gravitational force must be proportional to mass. As we will show later Newton was even able to obtain an approximate value of the universal constant G. Its precise determination will be described in the next chapter. The numerical value is

$$G = 6.67 \times 10^{-8} \text{ dyne cm}^2/\text{g}^2.$$

4.2.2 The Electromagnetic Interaction

Mass is only one of the many properties of matter. According to the molecular - atomic hypothesis to be discussed later, and the wide range of observational phenomena which are compatible with this hypothesis, all matter consists of a combination of a few kinds of atoms, each of which has a definite internal structure. The basic entities involved in the overall structure of the atom are two types of particles, an atomic nucleus and a number of electrons. Both the nucleus and the electrons possess the property of inertial mass. But in addition they possess a property which we describe as electric charge. It is the interaction of these charges with each other which we call the electromagnetic interaction. It is this interaction which is responsible for the formation of atoms and molecules. The electrons move in the atom around the nucleus under the influence of forces which constitute the electromagnetic interaction to form a stable system in which the attractive forces of the electric interaction are precisely balanced by the inertial forces as given by Newton's second law of motion. Since the particles also possess mass, there are gravitational forces as well, but the magnitude of the latter are very much smaller than the former and they play a negligible role whenever electromagnetic forces are present.

It is at this point that we note an important qualitative difference between electromagnetic and gravitational forces. The latter is always attractive whereas both attractive and repulsive forces exist between electric charges. Two identical charges always repel each other, but pairs of charges such as the electron and the atomic nucleus, or the electron and a proton, attract. Thus, one must conclude from the outset that there are two kinds of electric charge which we call positive and negative charge. According to the convention which has been adopted the charge on the electron is negative and the charge on the proton, and on the atomic nucleus, is positive.

Experiment shows that the negative charge on the electron and the positive charge on the proton are precisely equal in magnitude. Also the charge on an atomic nucleus is precisely an integral multiple of the protonic charge. It is this exact equality of the basic units of negative and positive charge that accounts for the fact that normal matter is electrically neutral. The number of electrons in each normal atom is equal to the number of basic units of positive charge on its nucleus, and these electrically neutral atoms are combined to form neutral molecules, and the latter combine to form matter as we know it. Consequently, there is normally on the macroscopic scale no manifestation of an electromagnetic interaction since the interacting bodies are normally uncharged. The only interaction between such bodies is gravitational. But within the atom, and for reasons to be discussed later, between atoms and molecules separated by a short distance there is an electromagnetic interaction. The intra-atom electromagnetic interaction is responsible for the stability of atoms. The interatom and intermolecular short range electromagnetic interaction is responsible for the cohesive properties of matter. As we will find, these interactions are all so enormously large compared to the gravitational interaction which is always present that the latter is completely negligible in its effects whenever the former are present. In fact, it was only very recently that the gravitational force on a single free electron due to the attraction of the earth was observed.

Electric charges exert forces on each other which depend on the magnitude and sign of the charges, on the distance of separation and on their relative motion. The simplest case is that of two charges which are at a fixed distance r from each other. If the two charges have the values q_1 and q_2 the force of attraction or repulsion is proportional to the product $q_1 q_2$ and to $1/r^2$. It is convenient to define the unit of charge in such a way that the proportionality factor is unity. Then in one set of units which we call the cgs electrostatic system of units

$$F = \frac{q_1 q_2}{r^2}.$$

This is to be interpreted as a repulsive force if the sign of $q_1 q_2$ is positive, otherwise it is attractive. It is the electrostatic part of the electromagnetic interaction. Like the gravitational force its magnitude varies as the inverse square of the distance of separation between the particles.

Additional contributions to the electromagnetic interaction occur if the charges are in motion. A more complete description of the electromagnetic interaction, including also a description of typical electric and magnetic phenomena on the basis of which the form of the interaction is determined, will be given in the chapter on electric and magnetic fields.

4.2.3 The Nuclear or Strong Interaction

The existence of this interaction was first suspected in 1932 when the neutron was discovered and it was realized that the atomic nucleus was composed of neutrons and protons. As we have already noted, the protons each possess one unit of positive electric charge. The neutrons have the same mass as the proton, but they possess no electric charge. Consequently, no attractive electrostatic interactions are present in the nucleus. In order to account for the tight binding of neutrons and protons to form a stable nucleus it was necessary to invent a new force. Certain properties which this force must have were evident from the beginning. The new force must be attractive and it must be essentially the same between neutron pairs, proton pairs, and between neutrons and protons. The range of the force, i.e., the separation distance below which the magnitude of the force is not negligible, must be of about the same dimensions as the nucleus or somewhat less, and within this range the force must be sufficiently strong to overcome the proton-proton electrostatic repulsion and in addition provide the strong binding which is observed for each of the constituent particles in the nucleus. Quantitative details of the nuclear interaction will be given in the chapter on the atomic nucleus.

4.2.4 The Weak Interaction

This interaction is mentioned here only for completeness; it is one of the four basic interactions. Its role, however, is not well understood. It is the interaction which is responsible for certain particle transformations. Thus, for example, a free neutron may change into a proton by emitting an electron and a lighter particle called the neutrino. Similarly, certain heavy nuclei may change into other nuclei by this and an analogous process. We speak of such processes as particle decays and nuclear decays. It is a kind of tearing-down process, and the forces which are responsible may be regarded in a certain sense as disruptive forces. Consequently, the weak interaction forces which are responsible for these processes are quite different from the other forces, each of which in its own way is responsible for the stability which we observe in nature. The weak interaction is of quite a different character and its role in the scheme of natural phenomena is a mystery. This interaction will not be discussed further in this course, although the decay phenomena for which it is responsible will be described qualitatively.

4.3 EMPIRICAL FORCE LAWS

In many physical situations of interest motion takes place under conditions which are not completely determined. For example, an object, say an airplane, moves through the air at high speed. The air which supports the plane against the downward pull of gravity also offers resistance to its forward motion. As the airplane moves, the air is compressed in some regions near the airplane, rarified in others, and always in motion. Some of the air flows smoothly around the airplane, but much of it does not. There is turbulence and there are random variations in the lines of flow. Clearly the situation is physically very complicated. However well one might understand the basic principles of air flow and the movement of rigid objects through a fluid medium, the problem of calculating all the details of motion in a situation such as this would be enormously complicated.

Let us consider another kind of physical situation. When two objects are in contact a force is required to move one with respect to the other. Consider for example an object resting on an inclined plane with no external forces acting other than the force of gravity. As the angle θ which the plane makes with the horizontal is increased, starting from zero, the component of gravitational force parallel to the plane, mg sin θ, increases. But the object does not start to move until θ is increased to some critical value. The failure of the object to be accelerated at angles less than this critical value is explained by the assertion that a frictional force is present. At sufficiently small values of θ this frictional force adjusts itself so that it exactly balances the downward force mg sin θ, and the net force is zero. A critical angle is reached because the frictional force can not exceed a certain maximum value. We now ask about the determination of the maximum frictional force from first principles. Can we, for example, use whatever knowledge we have of the basic interactions to calculate the maximum frictional force which can be sustained by two surfaces in contact? A little consideration based, let us say, on some further knowledge of the nature of surfaces shows that we can not. In general we are not dealing with surfaces that are perfectly smooth, and we are hardly ever dealing with surfaces which are completely free of contaminants. Even if we go to considerable effort to make sure that the surfaces are smooth and that they are clean, there will still be oxides and absorbed gases present. The actual magnitude of the frictional force will depend on all of these details, and obviously a calculation of the frictional force starting from the basic interactions could not lead to the correct results in the absence of detailed knowledge of the surface conditions. We can go further in our characterization of this type of problem. We can in fact achieve almost perfect smoothness and cleanliness of two surfaces. We can also remove all or most of the absorbed gases. Suppose now that we are dealing with two materials of the same kind, say two pieces of copper. The surfaces which we bring into contact are almost perfectly smooth, they are clean, and there are no absorbed gases upon them. We can go quite far experimentally in realizing these ideal conditions. When we do we find that the two pieces stick together. The atoms on the surface of one piece of material find themselves about as close to some atoms in the other piece as they are to some atoms in the piece to which they belong. Thus, ideally there is no frictional force between the surfaces of different bodies; there is only cohesion. In the absence of cohesion one has friction, but the magnitude of the friction depends on such noncalculable factors as roughness, contamination, and adsorbed gas.

Let us consider still another example. If force is applied along the length of a straight wire, the wire is changed in length. The magnitude of the change in length depends on the force applied. If the materials of which the wire is made are known and homogeneous and if the basic interactions are fully understood, the relation between applied force and change in length can be calculated. But the problem is obviously not a simple one, and in the end one may find that one understands problems of this type only in principle. To some extent it is the essence of physics to supply answers to problems of this kind only in principle. One can describe all of the connections between the basic interactions, the arrangement of atoms in the matter under consideration, the state of dynamical motion of the atoms within the crystal lattice, and still not be able, or even find it desirable, to give precise numerical answers to questions of the type we are now asking.

In each of the three examples we have cited the question of interest is concerned with a force law: what is the resistance of the air to a high speed airplane and how does the resistive force depend on the relevant parameters of the problem; what is the force of friction between two surfaces which are pressed together and on what factors does it depend; what is the force required to stretch a wire a given amount? In these and in many other examples of a similar character we find it desirable to determine the forces empirically. The force laws determine in this way then provide the basis for the further consideration of many interesting types of motion. They also provide the experimental data with which the results of a theory of each of the force laws based on first principles can be compared.

The determination of force as a function of the parameters in any given physical situation does not present any conceptual problems, though the actual execution of an experiment in which a force is measured may not always be easy. In principle one needs a force scale. One of the simplest is a spring whose changes in length as a consequence of the application of a force are accurately reproducible. Such a spring may have its various elongation calibrated using known forces, as for example, by comparison with the force of gravity. Another convenient device is a column of liquid in an open or closed tube. Such a column exerts a pressure (force per unit area), which is equal to the height of the column multiplied by the density of the liquid and the acceleration of gravity. Account must be taken also of pressure on the top surface of the liquid column, but this can be reduced to a negligible value.

The following examples of empirical force laws are typical of numerous phenomena to be found in nature. Each example belongs to a wider range of phenomena than is indicated by the brief description given.

4.3.1 Sliding Friction

The force required to move two bodies in contact along the surface of contact depends on the nature of the surfaces and on the force which is pressing the two bodies together. We will speak of this force as the maximum frictional force and denote it by the symbol F_r. Tangential forces smaller than F_r produce no motion. Tangential forces greater than F_r produce motion with acceleration. Since the force pressing two surfaces together is a normal force (perpendicular to the surfaces in contact), we denote it by N. The experimental result is that

$$F_r = \mu N,$$

where μ is a constant called the coefficient of friction. This coefficient depends on the nature of the surfaces, and it may depend also on the temperature and the atmospheric humidity. For a given roughness of surfaces it depends very much on the state of contamination. It does not depend on the area of contact between the two bodies.

In many cases one finds that the frictional force decreases after the motion has started, and it may in fact depend slightly on the velocity. Consequently one speaks of a coefficient μ_s of static friction and a coefficient μ_k of kinetic friction. The difference between μ_s and μ_k is negligible in some cases, as for example, in the contact of two dry metals, or for a material like teflon on metal. But for glass on glass μ_k is often less than half of μ_s and for many surfaces μ_k is of the order of 20% less than μ_s.

It should be noted that $F_r = \mu N$ is not a vector relation: F_r is a tangential force and N is a normal force. The equation gives a relation among magnitudes.

4.3.2 Viscosity of A Liquid or Gas

When the layers of a fluid medium move at different velocities a force may be found which is responsible for the conditions. Thus, in the flow of fluids through pipes, the fluid immediately adjacent to the walls is stationary and the fluid in the center is moving at the greatest velocity. In between the wall and the center of the pipe the various layers are moving at intermediate velocities with a continuous variation from zero velocity at the walls to the maximum velocity at the center. In order to maintain this condition the pressure on the fluid and within the fluid decreases in the direction of flow.

This kind of condition is most easily visualized by considering the flow of fluid between two horizontal planes. Let the lower boundary plane be at zero velocity and the upper plane at velocity v. Between the two planes the velocity of the fluid changes uniformly with distance from zero at the bottom to the maximum value at the top. In order to maintain this condition one finds that there must be a force on the two planes, the one at the top being in the direction of the fluid velocity, and the one at the bottom oppositely directed and equal in magnitude. This force is proportional to the area of the boundary planes, to the velocity v, and inversely proportional to the separation distance d between the planes. Denoting the force per unit area by F, the expression for F consistent with experiment is

$$F = \eta \frac{v}{d}.$$

We call η the viscosity coefficient of the fluid. It may be determined empirically by experiments based on the arguments which have just been given, but for experimental convenience the conditions need not be precisely those in terms of which we have described the force law. The viscosity coefficient is an important property of the fluid medium. For any given fluid consisting of known molecules interacting with each other in known ways, η may be calculated from first principles.

4.3.3 Resistance to Motion Through A Fluid

Any object moving through a fluid medium, e.g., the air or water, always experiences a resistive force, that is, a force which is opposite in sign to the direction of motion. For small bodies moving at low velocities this force is proportional to the velocity and the magnitude depends on the linear dimensions. It is convenient to express the experimental result in the form

$$F_r = k\ell v,$$

where k is a constant depending on the shape of the body and on the viscosity of the fluid medium, ℓ is a linear dimension of the object, and v is the velocity of motion.

If the object in question is a small sphere (of diameter less than about 1 mm in the case of air), conditions of fluid flow around the object are such that a precise calculation in terms of properties of the medium can be made. The result is known as Stoke's Law. It is

$$F_r = 6\pi\eta rv,$$

where r is the radius of the sphere and η is the viscosity of the medium.

For larger spheres and for spheres moving more rapidly a departure from the linear dependence on velocity is observed. A contribution to F_r proportional to v^2 begins to be observed and for higher velocities this contribution dominates. Even for raindrops falling in the atmosphere there is some deviation from the linear dependence of F_r on v.

For objects like airplanes one has

$$F_r = Cv^2.$$

In spite of the great complexity of this system the force law is a simple one, and a reproducible determination of the constant C is possible. The force is however given by a square law rather than a linear dependence on v.

4.3.4 Elastic Deformations

All rigid bodies change their shape under the application of a force. Well-known examples are the stretching or compression of a helical spring by a force acting parallel to the axis of the helix, the bending of a beam by the application of a force transverse to the beam, and the elongation of a wire by the application of a force along the length of the wire. In each case the magnitude of the deformation produced can be measured as a change in a linear dimension. The magnitude of the change is found to be always proportional to the magnitude of the applied force providing that the elastic limit of the material of which the body is made is not exceeded. Elastic limits are defined in terms of the maximum forces which can be applied without producing a permanent deformation in the shape of the body. Some materials such as steel have high elastic limits, and bodies made of such materials are called elastic. Other materials such as lead and rubber have small elastic limits, and for some others such as putty and dough the limit is effectively zero. Bodies made of materials of low elastic limit are described as inelastic.

Consider now an elastic body. To be specific let us take the case of a helical spring. In the absence of any applied force it has some length x_0. We might call this the equilibrium length. If we now apply a force F the length changes to $x_0 + x$, and if we apply a force 2F the length changes to $x_0 + 2x$, i.e., the change in length is proportional to the applied force. The force law may be expressed in the form $F \sim x$. It is convenient for many purposes to write the force law in terms of the opposing force in the spring. This force is the negative of the applied force. Introducing also a proportionality constant k we write the force law for the sping in the form

$$F = -kx.$$

The constant k is called the force constant. Its magnitude is determined by the size and shape of the body, by the materials of which it is made, and by the nature of the deformation which is being considered. When applied to any simple elastic deformation as the stretching of a helical spring or the bending of a flat spring, the force law is known as Hooke's law. The important property is the proportionality to displacement and the opposition of displacement and force. We note particularly that the force in the spring is a restoring force whose magnitude is directly proportional to the displacement from equilibrium.

Another example is the stretching of a wire by oppositely directed forces parallel to the axis of the wire. In this case it is convenient to use a generally valid definition of an elastic coefficient as a ratio of stress to strain and to define each of the latter in such a way that the coefficient is a pure material constant. Thus, if the wire has the equilibrium length ℓ and the cross-sectional area A, and if an applied force F produces an increase in length $\Delta\ell$, we define

$$\text{stress} = \frac{F}{A}$$

$$\text{strain} = \frac{\Delta\ell}{\ell}$$

and the elastic coefficient which we denote by Y as

$$Y = \frac{\text{stress}}{\text{strain}} = \frac{F\ell}{A\Delta\ell} \,.$$

A little consideration of this relation shows that Y depends only on the material of which the wire is made. It is called Young's modulus.

Other definitions of elastic coefficients can be given when they are needed. In all cases the strain is proportional to the stress for sufficiently small stresses and the elastic coefficient is defined simply as the ratio of these two quantities.

4.3.5 Surface Tension

The surfaces of liquids are always under tension. It is this tension which is responsible for the spherical shape of rain drops, the floating of a needle on the surface of water under certain conditions, and the rising of certain liquids in capillary tubes; e.g., the transport of fluids from the ground to the upper portions of growing plants.

The ultimate source of surface tension is the short-range attraction of the molecules in a liquid for each other. In the body of the fluid the attractive forces are equally strong in all directions. But near the surface the molecules are pulled predominantly toward the main body of the liquid and away from the surface. The net result of this unbalanced attraction is to produce forces which lead to a minimization of the surface area. These forces have the kinds of consequences listed above.

The magnitude of the unbalanced surface forces can be measured by producing free surfaces as in a soap film. If the film is made rectangular in shape the opposing forces on two opposite sides of the rectangle required to maintain the film can be measured. The result is that the force in the surface is simply proportional to the length of the surface edge. The proportionality factor is defined as the surface tension.

4.3.6 Molecular Forces

A somewhat different empirical force law than any we have discussed so far is the short-range force of molecules for each other. Since molecules are rather complicated structures, the calculation of the interaction between two molecules starting with the basic interactions (the electrostatic interaction in this case), is not necessarily a simple problem. The general features of this interaction are easily determined, but the quantitative details are not. Consequently, it is of some interest to have an empirical determination.

The result is that the force between molecules is attractive at long distances and repulsive at short distances. At an intermediate distance the force changes from attractive to repulsive, and consequently there is an equilibrium distance, the average distance of separation of the molecules in the bulk material, at which the force between molecules is zero. The attractive force is really not long range since its magnitude varies the inverse seventh power of the distance, and therefore it decreases to zero very quickly after reaching its maximum value just outside the equilibrium distance.

These molecular forces are responsible for a number of the macroscopic properties which have already been mentioned. They are responsible, for example, for the sliding frictional force which is present when two surfaces are in contact. Foreign molecules are generally present at such surfaces. As the two surfaces are moved with respect to each other, molecular attractions between molecules which are in motion with respect to each other generate atomic motions and set up vibrational waves in the adjacent medium which can carry away energy. If the foreign molecules are absent there may be adhesion, and this too is caused by the molecular interaction. Finally the molecular interaction can be used to explain Hooke's law of elasticity in bulk materials. Any deformation of the body produces a change in the average separation distance of the molecules along the line in which the deformation occurs. One can show that for small changes the molecular force changes between positive and negative values uniformly with distance. This is just the behavior required to explain the over-all force law.

5 MOTION UNDER THE INFLUENCE OF GRAVITATIONAL FORCES

5.1 INTRODUCTION

As described in section 4.2, all matter attracts all other matter in the universe with a force which has its origin in a property of mass. This is a very weak interaction in comparison with other forces to be studied in later chapters, but for matter as observed in the large-scale universe it is usually the only force. It is the force which determines the motion of the planets, comets, and satellites, and at the same time it is the force which is responsible for the free fall of objects at the surface of the earth, and for the property of materials which in everyday language we describe as weight. Since it is possible in laboratory experiments to eliminate the effects of other forces, the range of conditions under which the gravitational force can be studied is a very wide one, extending, insofar as the magnitude of the interacting masses are concerned, from a few grams to masses of the order of sun's masses, and, insofar as separation distances are concerned, from a few centimeters in a laboratory experiment to the many millions of miles separating the planets and their attracting center in a solar system.

The great achievement of Newton, among many others, was to recognize that this wide range of phenomena could be considered together and described in terms of a few simple basic laws of physics which he then proceeded to formulate. We have already considered the basis for his formulation of the equations of motion. In the section on basic interactions we have also described some aspects of the basis for his formulation of the laws of universal gravitation. But there were also other stimuli and other clues. Of major importance in this connection were the results of the life time work of Johannes Kepler (1571-1630), student of Tycho Brahe and immediate predecessor to Newton. Using the very extensive and accurate astronomical observations of Brahe, Kepler succeeded in synthesizing all astronomical data in the form of three general laws of planetary motion. For Newton these laws provided both a clue to the nature of the interacting forces and at the same time a severe test of any hypothesis which might be advanced to explain the motion. Kepler's laws of planetary motion will be described in the next section.

As the student will have noted, the law of universal gravitation as stated in the previous chapter can apply only to bodies whose extensions in space are negligible in comparison to the distance of separation. In general it is only for such cases that there is a definite separation distance r. Strictly speaking, the law of universal gravitation applies only to the interaction of two mass elements Δm_1 and Δm_2 in two volume elements ΔV_1 and ΔV_2, the linear dimensions of each of which are negligible in comparison to the separation distance. Newton was keenly aware that the law could not be applied without further consideration to such bodies as the earth and the moon, or the earth and a baseball, the radii of one or both of which are comparable, or at least, not negligible in comparison with the separation distance of the two bodies. It was his failure to solve this problem in the early stages of his consideration of the law of universal gravitation which caused him to withhold the announcement of the great success which he had in fact achieved. Many years before the actual publication of these results he was able to give convincing arguments in support of his conclusion that a single type of force was capable of explaining both the astronomical and terrestrial motions, that the direction of the force was along the lines of centers connecting two spherically shaped bodies, and that the magnitude of the force was inversely proportional to the square of the distance between their centers. But the statement that "the law of gravitational interaction between two large spheres in which the distribution of matter is spherically symmetrical (as is presumably the case for all the planetary bodies) is given by the same expression as for two point masses, provided only that the centers of the spheres are used as the points between which the separation distance is measured," requires the use of the calculus for its proof. Newton invented the calculus, and he succeeded eventually in proving the statement which we have just given. In our subsequent consideration we will assume that the expression

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

applies both to point particles of masses m_1 and m_2 and to spherical distributions of matter in which the total masses are m_1 and m_2 and the distance between centers is r.

5.2 KEPLER'S LAWS

Kepler's three laws of planetary motion are concerned with the nature of the orbits, the speed of a planet at various points in a given orbit, and the way in which the periods of motion change as one goes from one orbit to another in a given planetary system. Since no one up to this time, including Kepler, had any notion that the motion of a heavenly body could be anything but circular, or a superposition of circular motions, Kepler's stunning conclusion (after many years of futile effort with circular orbits), that the orbit was in fact a simple mathematical curve of which the circle was only a special case represented a great break with the past. Since he found that the orbits were not in general circular, there was no longer any basis for thinking that the speed of a planet in a given orbit was constant. Thus, a statement with respect to the varying speed of motion, a kind of variation which Kepler found could be expressed with the utmost simplicity, became the second law. Finally it was natural to look for a unifying principle relating all orbits in a single planetary system. This was not easy to find and it required nine more years of searching and a considerable faith in the basic unity and harmony of nature, a faith which Kepler surely had, in order to persist so long. Finally he found the general principle which became his third law. The student interested in further details of the historical development will find a summary beginning with the astronomy of ancient Greece in Holton, Introduction to Physical Science, chaps. 6-11 (see footnote 2).

We will now state and briefly describe each of the three laws of Kepler.

5.2.1 Kepler's First Law

This law states that the planets move in elliptical paths, with the attracting center, the sun in the case of the solar system, at one of the two foci of the ellipse.

The properties of the ellipse have been known since the second century B.C. Like the circle it may be described by a simple equation in rectangular coordinates:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

where a and b are called the half major and half minor axes ($a > b$). For $a = b$ the equation describes a circle of radius a. Thus, the circle is a special case of the ellipse and in accordance with Kepler's first law, a circle is a possible planetary orbit.

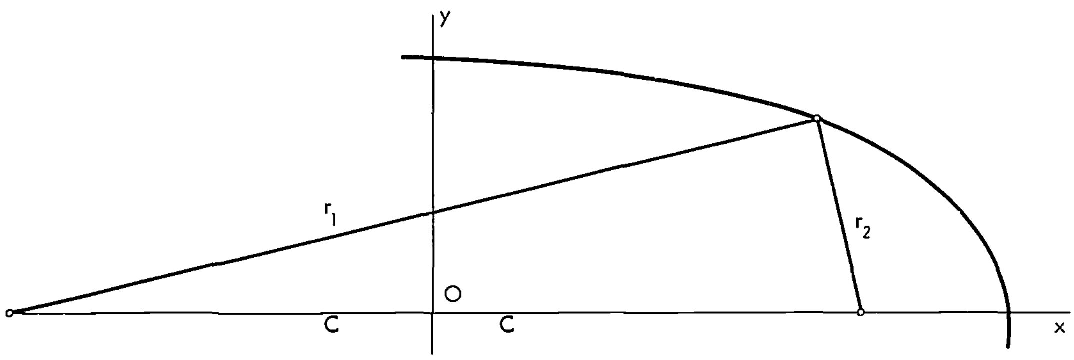

Fig. 5.1

Other properties of the ellipse which are found useful in an elementary treatment are the following:

(a) The ellipse is a curve drawn in such a way that the sum of the distances to two fixed points, $r_1 + r_2$ in Fig. 5.1, is a constant. The two fixed points are called the foci of the ellipse. With a rectangular coordinate system oriented as shown in the figure the major axis 2a is along x and the minor axis 2b is along y.

(b) The constant distance $r_1 + r_2$ is equal to the major axis 2a. This is evident from a consideration of the point on the ellipse for which r_1 has its maximum value (and r_2 its minimum value).

(c) The foci are separated by the distance 2c where

$$c = \sqrt{a^2 - b^2}.$$

This follows from a consideration of the point on the ellipse for which $r_1 = r_2$. At this point

$$r_1{}^2 = r_2{}^2 = b^2 + c^2$$

$$2a = r_1 + r_2 = 2r_1 = 2\sqrt{b^2 + c^2}$$

$$a = \sqrt{b^2 + c^2}$$

(d) The departure of an ellipse from a circle of radius a is conveniently described in terms of an eccentricity parameter e defined by

$$e = \sqrt{1 - \frac{b^2}{a^2}} = \frac{c}{a}.$$

The separation of the two foci is 2c = 2ea, and for a circle e and c are equal to zero.

(e) Another expression for the ellipse is in terms of the distance r to one of the two foci, say the one on the right, and the angle θ which the line connecting the planet to this focus makes with the positive x axis. This is

$$\frac{1}{r} = \frac{a}{b^2}(1 + e \cos \theta).$$

This expression for the ellipse may be obtained by starting with

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

substituting

$$x = c + r \cos \theta$$

$$y = r \sin \theta,$$

using the previously determined relations to eliminate c and introduce e, and carrying out the required algebraic manipulations.

5.2.2 Kepler's Second Law

This law states that the velocity of a planet in its elliptical orbit is such that a line drawn from the attracting center at one focus of the ellipse to the planet sweeps out equal areas in equal intervals of time. This is called the law of constant areal velocity.

The restatement of this law in terms of linear velocities will be described when we consider Newton's derivation of the empirical laws from the equations of motion and the force law. For the present let us consider only the linear velocities at two points on the ellipse which are respectively the two points of minimum and maximum approach to the attracting center. These two points lie at the extremities of the major axis and are called the perihelion and aphelion, respectively. At these points, and only at these points, the velocity of the planet is perpendicular to the line drawn from the planet to the attracting center. Denote the lengths of these lines as r_a and r_p. If the corresponding velocities are v_a and v_p the areas swept out in unit time are $\frac{1}{2}r_a v_a$ and $\frac{1}{2}r_p v_p$. But according to Kepler's second law these areas are equal. Then

$$\frac{v_a}{v_p} = \frac{r_p}{r_a}.$$

5.2.3 Kepler's Third Law

This law describes the manner in which the period of a planet in its orbit varies as one goes from an orbit of one size to another, all within a given planetary system. The law says that the square of the period is proportional to the cube of the mean radius. One can show by a simple geometric argument based on the statement $r_1 + r_2 = 2a$ that the mean radius is equal to a. The third law can then be expressed in the form

$$T^2 = ka^3,$$

where k is the same for all planets in a given system.

5.3 DERIVATION OF KEPLER'S LAWS FROM THE LAW OF UNIVERSAL GRAVITATION

We will discuss Kepler's laws ɔm the point of view of their being a consequence of the law of universal gravitation. This procedure is the inverse of the historical development. As we have indicated, Newton probably rested heavily on his knowledge of Kepler's laws for inspiration and guidance in his search for a physical interpretation of planetary motion. Thus, the mathematical form of the orbit in which the planet moves, as described by Kepler, reinforced his conviction that the force of attraction between two masses falls off as the inverse square of the distance between them, and the constancy of the areal velocity told him that the force must be directed along the line of centers connecting the two bodies. These conclusions are not immediately obvious to the mathematically uninitiated student, but after some reflection they were obvious to Newton, and we will be able to demonstrate important aspects of them here. Our procedure, however, will be to start with the force law, and then discuss Kepler's laws as an immediate consequence, basing the proof of course on the equations of motion.

We begin with Kepler's second law rather than the first. We will now show that the constancy of areal velocity is an immediate consequence of all force laws that have the property that the force is directed along the line of centers connecting two bodies. Such forces are called central forces. We encounter forces in nature which are not central, and for these the property of constant areal velocity does not exist. But other forces besides the force of gravitational attraction are central and for all of these the motion is such that the areal velocity is constant. Thus, Kepler's second law falls into a very general category. We will proceed now to prove it in this general sense. We will prove that the areal velocity is constant for any force which is central. We will then know it to be true in the special case of a gravitational force.

Consider three nearby points in

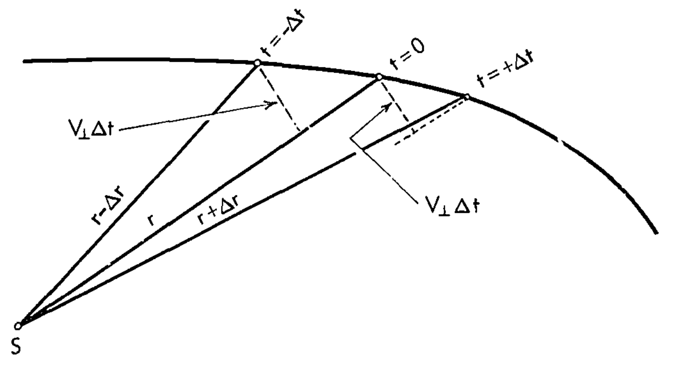

Fig. 5.2

the orbit which are separated by two small and equal time intervals Δt. For convenience we will take t = 0 at the midpoint and t = − Δt and t = + Δt at the earlier and later points, respectively. Draw lines from the attracting center at S to the planet at each of the three points, and denote by r − Δr, r, and r + Δr the changing magnitude of the distance between the two bodies, (see Fig. 5.2). We note also that the velocity is changing as the planet moves along the orbit. Denote by v the velocity t = 0. We resolve this velocity into two components, one, $v_\perp$ which is perpendicular to r, and the other, $v_\parallel$ which is parallel to r. This brings us to the critical point in the argument. Since the force is central it is along the direction r. By Newton's second law of motion this force can change only the parallel component. Then $v_\perp$ is a constant. It will change with r as a consequence of motion to another part of the orbit, but at a given r, $v_\perp$ is not changing as a consequence of the force. Then in the diagram the dotted lines drawn perpendicular to r are of equal lengths $v_\perp\Delta t$. Consider now the two triangles representing the areas swept out by r in moving from t = −Δt to t = 0, and from t = 0 to t = +Δt. These triangles have the common base of length r and they both have the altitude $v_\perp\Delta t$. Then the area of both triangles is $\frac{1}{2} rv_\perp\Delta t$, and the areal velocity in both intervals is $\frac{1}{2}rv_\perp$. The areal velocity is a constant because $v_\perp$ is not changed by the central force. As the planet moves around the orbit, r changes and therefore, $v_\perp$ changes, but the areal velocity

$$v_A = \frac{1}{2} rv_\perp,$$

is a constant. Thus, Kepler's second law is proven for central forces. An equivalent statement is that the angular momentum A about the attracting center S is a constant. From section 3.6, $A = mrv_\perp = 2mv_A$, where m is the mass of the planet. Since v_A is constant, A is constant. The latter also follows directly from the relation proved in Chapter 3 that torque is equal to time rate of change of angular momentum. Since for a central force the torque is zero, it follows that the angular momentum is a constant.

We turn now to the proof of Kepler's first law. In order to prove it we must show that the orbit has unique properties which we recognize as belonging to the mathematical curve which is called the ellipse. Some of the properties of this curve have already been described. There are several methods which can be used to relate these mathematical properties to characteristics of the motion as determined by the form of the gravitational attraction. One method is to start with the equation of motion. This method requires a discussion in terms of instantaneous rates of change and is consequently a method making use of the calculus. Another method is purely geometrical, but the geometrical argument is rather long and tedius. A third method is based on the consideration that the total energy $E = K + U$ is a constant. Both the kinetic energy K and the potential energy U change as the planet moves around the orbit but the sum is unchanged. Using the result already demonstrated in the proof of the second law that changes of velocity due to the force take place only along the line of centers, we can obtain an expression for the velocity, and therefore the kinetic energy K, as a function of position along the orbit in terms of the angle which the line of centers makes with some fixed line. We also need to have an expression for the potential energy of the two bodies as a function of the separation distance. The potential energy, and a new concept, the potential, are discussed in section 5.5. The derivation of the first law follows. But since this derivation is not essential to the student's understanding of the remainder of the text, the derivation will be found in an Appendix rather than in this chapter. We will note here only that one obtains by this procedure an expression for the separation distance r in terms of position along the orbit which is precisely of the form which we recognize as that of an ellipse. Thus, Kepler's first law is proven. Essential to the proof are both the central character of the force and its variation as the inverse square of the distance.

We will prove the third law only for the special case of motion in a circle. Consider a planet of mass m moving around a sun of mass M at a constant fixed distance of separation R. According to the law of universal gravitation, a force

$$f = G\frac{Mm}{R^2}$$

acts on the planet. By Newton's third law of motion an equal and opposite force acts on the sun. Thus, both bodies are accelerated toward each other, that of the planet being equal to f/m, and that of the sun f/M. We consider here the case that M is very large in comparison with m so that the acceleration of the sun in comparison with that of the planet is negligible. Under these circumstances, we can regard the sun as fixed in space. If we simplify the problem still further by choosing as the elliptical path in which the planet moves the special case of a circle in which the speed v of the planet is a constant, then the acceleration of the planet is v^2/R where R, the distance to the sun, is also the radius of the circle in which the planet moves. Setting this acceleration equal to f/m we obtain

$$v^2 = \frac{GM}{R}.$$

This result may be expressed as a relation between the period T and R. Since $vT = 2\pi R$ the result is

$$T^2 = \frac{4\pi^2 R^3}{GM}.$$

This is the relation we set out to derive. For the more general case of an ellipse of major axis 2a the relation is

$$T^2 = \frac{4\pi^2 a^3}{GM}.$$

The period does not depend on the eccentricity of the ellipse but only on its major axis. In accordance with the statement of the third law $T^2 \sim a^3$. The derivation gives us in addition the magnitude of the proportionality factor, $4\pi^2/GM$.

5.4 MEASUREMENT OF G

Since G is a universal constant, its numerical value is of great importance. However, it is a difficult quantity to measure with precision and more than a century passed after Newton's enunciation of the law of universal gravitation before an experiment especially designed to measure G was successfully performed. Newton was able to provide however an estimate of its magnitude. We will describe two methods, both involving the mass of the earth. Since the mass of the earth cannot be known until G has been determined, and therefore had to be estimated or guessed, the value of G obtained by these methods was not very reliable.

One method is based on observations of the moon, an earth satellite. The moon's orbit is nearly circular and the relation derived in the proof of Kepler's third law,

$$T^2 = \frac{4\pi^2 R^3}{GM},$$

is a valid one. In this case M is the mass of the earth since it is the massive center around which the planetary motion is taking place, R is the distance from the earth to the moon and T is the period of the moon. The mass of the earth may be calculated from the relation

$$M = \frac{4}{3}\pi R_E^3 \overline{\rho},$$

where R_E is the radius of the earth and $\overline{\rho}$ is the average density. Since the latter is not known in the absence of measurements to be described below any calculation of G based on an estimate of $\overline{\rho}$ is bound to be inaccurate.

A second method makes direct use of the law of universal gravitation as a basis for calculating the force of gravity on objects near the earth's surface. The force on an object of mass m is

$$F = mg = G\frac{Mm}{R_E^2}$$

or

$$G = \frac{gR_E^2}{M} = \frac{3}{4\pi}\frac{g}{R_E\overline{\rho}}.$$

The acceleration of gravity g, and the radius of the earth R_E are known, but the average density $\overline{\rho}$ must again be estimated.

The third method involves a direct measurement of the force between two known masses at a known distance of separation. This is the Cavendish experiment performed by Henry Cavendish in 1797. In this experiment two small lead spheres are connected by a horizontal rod which is suspended at the center from a thin quartz fiber. Two other lead spheres are placed symmetrically with respect to the suspended spheres so that the distances of separation between pairs can be made quite small and the force of attraction acts to twist the quartz fiber. The force of attraction can be measured by observing the angle through which the suspension is turned and using a calibration of the angle in terms of known forces. The experiment is easily demonstrated and an accurate value of G can generally be obtained. As stated in Chapter 4, the accepted value is

$$G = 6.67 \times 10^{-8} \text{ dyne cm}^2/\text{m}^2.$$

Using this value of G in the two previous experiments one obtains a value for $\overline{\rho}$, the average density of the earth, and therefore the mass of the earth. The result is

$$\overline{\rho} = 5.5 \text{ g/cm}^3.$$

Thus, Cavendish referred to his experiment as an experiment in weighing the earth.

5.5 THE GRAVITATIONAL POTENTIAL

The change of potential energy of a system has been defined as the work which must be performed by an outside agent in order to change the configuration within the system. We wish now to calculate the work done when the separation distance between two masses m and M is changed from some value r to arbitrarily larger dis-

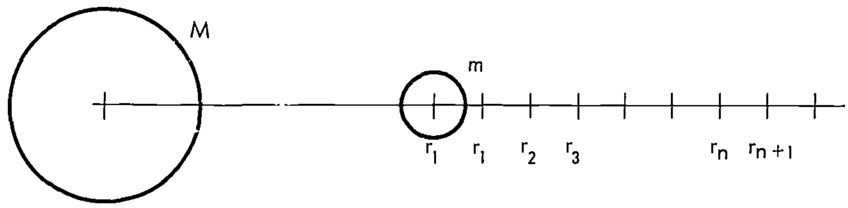

Fig. 5.3

tances. In order to calculate this work we will let the separation distance increase in small steps, first from r to r_1 as shown in Fig. 5.3, then from r_1 to r_2, then from r_2 to r_3, and so on. In each of the intervals the variation of the force with distance will be very small because the intervals are small. Consequently the work done may be calculated in each interval by using the average force in the interval and multiplying by the short distance through which the bodies are moved relative to each other. Let us calculate the work done when m is moved from the point marked r in the figure to r_1. At r the attractive force is

$$G\frac{mM}{r^2},$$

and at r_1 it is

$$G\frac{mM}{r_1{}^2}.$$

Since the quantity $1/rr_1$ lies between $1/r^2$ and $1/r_1{}^2$ in magnitude, it is plausible to assume that the average force in this interval is given by

$$\bar{f} = G\frac{mM}{rr_1}.$$

This is in fact the case as may be demonstrated rigorously using the calculus. We will simply assume that the expression is correct. Then the work done in this interval is

$$W = \bar{f}(r_1 - r) = GmM\left(\frac{1}{r} - \frac{1}{r_1}\right).$$

Similar expression will be obtained for the work done in succeeding intervals. Thus, between r_n and r_{n+1} the work done is

$$W = GmM\left(\frac{1}{r_n} - \frac{1}{r_{n+1}}\right).$$

The work done in any large interval is merely a sum of such terms. Starting with m at the point r the work done in moving m out to an arbitrarily distant point is accordingly

$$W = GmM\left[\left(\frac{1}{r} - \frac{1}{r_1}\right) + \left(\frac{1}{r_1} - \frac{1}{r_2}\right) + \left(\frac{1}{r_2} - \frac{1}{r_3}\right) + \ldots\right].$$

We observe that there is considerable cancellation of terms in this expression. In fact, only the end terms in any such summation over a finite number of intervals survive. If the summation is continued indefinitely the second of these two end terms becomes smaller and smaller. In the limit that we move from the point r to an infinite distance it goes to zero. Consequently, the work done in changing the separation distance from r to ∞ is given by the simple expression

$$W = \frac{GmM}{r}.$$

But this is the change of potential energy, and it is observed to be positive. If we now arbitrarily define the potential energy at infinity to be zero then the potential energy at r is

$$U = -\frac{GmM}{r}.$$

Using this result for potential energy of one body m in the gravitational field of M (or of M in the gravitational field of m) it is convenient to define the related concept of potential. Putting our attention on one of the two bodies, say M, we consider the potential energy of a body of unit mass in the gravitational field of M at the separation distance r. This potential energy is defined as the potential ϕ due to M. Thus,

$$\phi = -\frac{GM}{r}.$$

The potential energy U of a body of mass m is then obtained by multiplying ϕ by m, i.e.,

$$U = \phi m.$$

The concept of potential will be developed more fully in a succeeding chapter dealing with electrostatics. Since, however, the concept is a general one for a large class of fields, it is useful to introduce it at this point.

It is of some interest now to return to the concept of gravitational potential energy and to use the principle of energy conservation as a basis for calculating the escape velocity of any object from the vicinity of another. We may be interested for example, in knowing the velocity with which an object at the surface of the earth must be projected in order that it will escape to infinity. In order to simplify the problem the influence of the sun and all other planets in the solar system will be neglected for the purpose of the calculation. We simply observe that at the earth's surface the total energy of a body of mass m moving with a velocity v_0 is

$$E = K + U = \tfrac{1}{2}\, m v_0^2 - \frac{GmM_e}{R_e}$$

where M_e and R_e are the mass and radius of the earth, and use has been made of our knowledge of the potential energy, the energy at infinity being taken as zero. If we now wish the body to escape to infinity the total energy E can not be less than zero. If it is different from zero and positive it will be all kinetic. Thus, the condition that the body is barely able to escape is the condition that E = 0. Solving for v_0 we obtain

$$v_0^2 = \frac{2GM_e}{R_e}.$$

For numerical purposes in problems of this kind it is convenient to rewrite the product $G M_e$ by using the relation

$$mg = G\frac{mM_e}{R_e^2},$$

for the value of g at the earth's surface. This gives

$$GM_e = g\, R_e^2.$$

Then the escape velocity is

$$v_0 = \sqrt{2gR_e}$$
$$= 1.1 \times 10^6 \text{ cm/sec} = 7 \text{ miles/sec},$$

where we have used g = 980 cm/sec and $R_e = 6.37 \times 10^8$ cm.

5.6 OTHER APPLICATIONS

In the same way as observations on the period and radius of the moon may be used to measure the mass of the earth, once G has been determined, the mass of any body which is the center of a planetary system may be determined if sufficiently accurate data on its satellites are available. The mass is given by

$$M = \frac{4\pi^2 R_s^3}{GT_s^2},$$

where R_s and T_s are the orbital radius and period respectively of the satellite. In this way the mass of our sun can be determined from orbital data relating to the earth or any of the other planets. For orbits which are not circular, or approximately so, the quantity R_s must be interpreted as the half-major axis a_s. Similarly, the mass of Jupiter can be determined from observations on its moons. As is to be expected on the basis of internal consistency of the interpretation and the proven validity of Kepler's third law, different determinations of mass obtained from the data on different planets are in agreement.

The third law is useful also for establishing a direct comparison of two orbital periods when the ratio of the orbital radii are known. From the derived relation we obtain the ratio

$$\frac{T_1^{\,2}}{T_2^{\,2}} = \frac{R_1^{\,3}}{R_2^{\,3}},$$

where 1 and 2 refer to two different planets in the same planetary system. Applying this relation, for example, to the moon and an artificial earth satellite in an orbit approximately 100 miles above the earth's surface, one obtains for the satellite T_S = 0.061 days = 1.46 hours, using T_M = 27.3 days, R_S = 4100 miles and R_M = 240,000 miles.

Other applications may be discussed as problems assigned to the student.

Appendix A PROOF OF KEPLER'S FIRST LAW

The following proof is based on the principle of energy conservation together with a conclusion which may be drawn from the concept of the velocity circle. The latter concept must be developed and then used for the specified purpose. The over-all proof proceeds in three steps which will now be described. A fourth step then gives the dependence of total energy on the size of the orbit.

1. We first prove that as the planet moves in its orbit the change Δv in the magnitude of the velocity corresponding to a small change in position along the path is proportional to the change in angle $\Delta\theta$ of the line drawn from the planet to the sun.

A closed path is assumed from the outset. The attracting center is at S, and the position of the planet with respect to S is specified by r and θ with θ measured from the line drawn to S at the distance of closest approach, (see Fig. 5.4). Let this distance be denoted by r and the velocity at this point in the orbit by v_p. From Kepler's second law as proved in Chapter 5,

$$rv_{\perp} = r_p v_p .$$

Using

$$v_{\perp} = \frac{r\Delta\theta}{\Delta t}$$

$$r^2\Delta\theta = r_p v_p \Delta t .$$

From the force law

$$F = ma = m\frac{\Delta v}{\Delta t} = G\frac{mM}{r^2}$$

$$\Delta v = \frac{GM}{r^2}\Delta t .$$

Eliminating r^2 between these equations

$$\Delta v = \frac{GM}{r_p v_p}\Delta\theta .$$

The direction of the vector $\Delta\vec{v}$ is toward S.

2. The velocity circle.

Construct a velocity diagram by beginning with the vector $\vec{v}_p$ and adding successive increments $\Delta\vec{v}$ each proportional in magnitude to successive increments $\Delta\theta$. Since each $\Delta\vec{v}$ points toward S, the successive increments make angles with respect to each other which are equal to the change in θ. Then the increments fall within a velocity circle. The center of this circle C is displaced from the origin 0 of the velocity diagram if the velocity $\vec{v}_a$ at the point of maximum distance from S is not equal in magnitude to $\vec{v}_p$. The diagram which is illustrated (Fig. 5.5), shows a particular velocity $\vec{v}$ obtained after adding a certain number of increments $\Delta\vec{v}$ with a combined angular change θ. The radius of the velocity circle is

$$V = \tfrac{1}{2}(v_a + v_p),$$

and the distance between C and 0 is

$$\delta = V - v_a = \tfrac{1}{2}(v_p - v_a).$$

Then

$$v^2 = V^2 + \delta^2 + 2V\delta\cos\theta$$

$$= \tfrac{1}{2}(v_p^2 + v_a^2) + \tfrac{1}{2}(v_p^2 - v_a^2)\cos\theta .$$

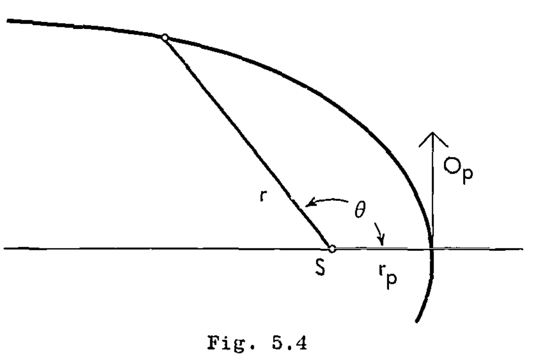

Fig. 5.4

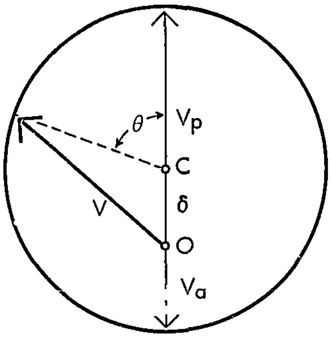

Fig. 5.5

3. Substitution into the energy equation

$$E = K + U = \tfrac{1}{2}mv^2 - G\frac{mM}{r},$$

where for U we are using the result derived in section 5.5, gives

$$\frac{1}{r} = \frac{1}{4GM}\left(v_p^2 + v_a^2 - \frac{4E}{m}\right)$$

$$+ \frac{v_p^2 - v_a^2}{4GM}\cos\theta.$$

From

$$E = \tfrac{1}{2}mv_p^2 - G\frac{mM}{r_p} = \tfrac{1}{2}mv_a^2 - G\frac{mM}{r_a}$$

$$v_p^2 + v_a^2 - \frac{4E}{m} = 2GM\left(\frac{1}{r_p} + \frac{1}{r_a}\right)$$

$$v_p^2 - v_a^2 = 2GM\left(\frac{1}{r_p} - \frac{1}{r_a}\right)$$

Then

$$\frac{1}{r} = \tfrac{1}{2}\left(\frac{1}{r_p} + \frac{1}{r_a}\right) + \tfrac{1}{2}\left(\frac{1}{r_p} - \frac{1}{r_a}\right)\cos\theta.$$

Comparison with one of the relations described in the discussion of Kepler's laws in Chapter 5 shows that the path is an ellipse.

4. The energy can be expressed in terms of velocities alone, or in terms of the size of the orbit

Use $r_a v_a = r_p v_p$ repeatedly starting with

$$Gm = \tfrac{1}{2}\frac{v_p^2 - v_a^2}{\frac{1}{r_p} - \frac{1}{r_a}}$$

$$\frac{GM}{r_p} = \tfrac{1}{2}\frac{v_p^2 - v_a^2}{1 - \frac{r_p}{r_a}} = \tfrac{1}{2}\frac{v_p^2 - v_a^2}{1 - \frac{v_a}{v_p}}$$

$$= \frac{v_p}{2}(v_p + v_a).$$

Then

$$E = \tfrac{1}{2}mv_p^2 - \frac{GmM}{r_p} = -\tfrac{1}{2}mv_p v_a.$$

Since also

$$GM = \tfrac{1}{2}v_p r_p(v_p + v_a) = \tfrac{1}{2}v_a v_p(r_a + r_p)$$

$$\tfrac{1}{2}v_p v_a = \frac{GM}{r_a + r_p}$$

$$E = -\frac{GmM}{r_a + r_p} = -\frac{GmM}{2a},$$

where 2a is the major axis.