CHAPTER

1

Introduction

Toward the end of the nineteenth century it seemed quite apparent to all physicists that the general concepts of what we now call "classical physics" were adequate to describe all physical phenomena. Classical mechanics, first formulated by Isaac Newton in the late seventeenth century, had by this time reached full bloom, and evidently provided a completely valid framework for the treatment of the dynamics of material bodies. Complementing classical mechanics was classical electrodynamics, finalized by James Clerk Maxwell in the latter half of the nineteenth century, which described all the properties of the electromagnetic field, and which in particular gave an intelligible account of the wave nature of light.

During the first quarter of the twentieth century, as physicists turned from their successful treatment of the macroscopic world to an examination of the microscopic world, a number of unexpected difficulties arose. Broadly speaking, these difficulties fell into two general categories.

First was the discovery of instances in nature in which certain physical variables assumed only *quantized* or *discrete* values, in contrast to the continuum of values expected on the basis of classical physics. For example, in order to explain the observed intensity spectrum of electromagnetic radiation inside a constant-temperature cavity (so-called "black-body radiation"), Max Planck in 1900 found it necessary to permit each atomic oscillator in the walls of the cavity to radiate energy only in the discrete amounts

$$h\nu,\ 2h\nu,\ 3h\nu,\ \ldots$$

Here, ν is the intrinsic frequency of the radiating oscillator (the cavity walls were assumed to contain oscillators of all frequencies), and h is a universal constant, now called *Planck's constant*, with the value

$$h = 6.625 \times 10^{-34} \text{ joule} \cdot \text{sec} \qquad (1\text{-}1)$$

As another example, in order to account for the spectrum of radiation emitted by excited hydrogen atoms, Niels Bohr in 1913 found it necessary to permit the angular momentum of the orbital electrons to have only the discrete values

$$h/2\pi, 2h/2\pi, 3h/2\pi, \ldots$$

There were several other instances of such "quantum effects" uncovered in the early part of the twentieth century. In each case, the quantization of the appropriate variable amounted to an *ad hoc* hypothesis, and was without precedent in earlier applications of classical physics.

The second category of difficulties which beset classical physics, as applied to the microscopic world, concerned the distinction between *waves* and *particles*. By 1900 it was generally believed that light was a wave, while the electron was a particle. However, concerning the nature of *light*, Albert Einstein in 1905 put forth his theory of the photoelectric effect, which indicated that a light beam of frequency ν behaves as though it were a collection of *particles*, each with an energy

$$\epsilon = h\nu$$

Einstein's hypothesis was a bold extrapolation of Planck's theory of blackbody radiation, but it was subsequently borne out in great detail by precise experimental investigations of the photoelectric effect; it received further dramatic support in 1923 when A. H. Compton showed that these light particles, called "photons," could actually be bounced off electrons according to the usual rules of classical mechanics. Meanwhile, concerning the nature of the *electron*, C. Davisson and L. Germer showed in 1927 that, by scattering a beam of electrons off a crystalline lattice of atoms, one could obtain diffraction patterns virtually identical to those which result from the crystal-scattering of X-rays. In fact, they showed that a beam of electrons of momentum p produced a diffraction pattern characteristic of a *wave* with wavelength

$$\lambda = h/p \tag{1-2}$$

in exact agreement with the conjecture made three years earlier by L. de Broglie. In short, light was found to behave sometimes as a particle and sometimes as a wave, and the electron was found to behave sometimes as a particle and sometimes as a wave! These results evidently implied some sort of "wave-particle duality" in nature which was quite unintelligible in terms of purely classical concepts.

It gradually became apparent during the first part of the twen-

tieth century that these two difficulties—namely the quantization of physical variables and the wave-particle paradox—bore the totally unexpected message that the *microscopic* world was simply not intelligible in the context of classical physics, and that a radically different approach was needed. As it happened, such a new approach was not long in coming: by 1930, through the efforts of W. Heisenberg, I. Schrödinger, M. Born, N. Bohr, P. A. M. Dirac, and many other physicists, a bold new system of mechanics called "quantum mechanics" had been devised.

The basic tenets of quantum mechanics are in many respects quite foreign to the concepts and attitudes of classical physics—so much so that there were and still are many eminent physicists who find some of these tenets philosophically unsatisfactory. Certainly it would be presumptuous to assert absolutely that quantum mechanics, as currently formulated, is the only or even the best possible way of understanding physical phenomena. However, there is no denying the fact that quantum mechanics, in its present form, has been amazingly successful from an "operational" point of view; that is, its predictions, no matter how unusual, have always been very much in accord with experimental observations. This, of course, is the reason for the acceptance of modern quantum theory by the overwhelming majority of physicists today.

It is our intent in this book to give a concise, simplified account of the main theoretical structure of quantum mechanics. To this end, we begin in Chapter 2 by presenting the "mathematical language" of quantum mechanics, assuming on the part of the reader mainly a reasonable grasp of elementary calculus. In Chapter 3 we review briefly the essential features of classical mechanics, in order that we may be able to readily compare and contrast the new with the old. In Chapter 4 we develop, under several simplifying restrictions, the basic formalism of quantum mechanics. Our method of presentation will be essentially "postulative-deductive"; that is, we shall lay out a number of postulates, and we shall try to derive, develop, and synthesize the implications of these postulates into a reasonably coherent theoretical framework. We shall emphasize neither the historical evolution of quantum mechanics, nor the applications of the theory to the solutions of various types of problems. Rather, we shall be chiefly concerned with understanding the *structure* and *spirit* of the theory itself. In particular, we shall try to see how quantum mechanics manages to subsume under a single, self-consistent point of view, the "common sense" of *macroscopic* physics along with the "obvious paradoxes" of *microscopic* physics. Following our development of the general theory, we shall consider

briefly its application to one simple idealized physical system, and this we do merely in order to illustrate various aspects of the theory. We conclude with a short discussion of how the removal of some of the simplifying restrictions which we imposed upon our development of quantum mechanics, can be expected to affect the overall theory.

There are a liberal number of exercises sprinkled in with the text. The majority of these exercises are not of the "problem-solving" variety; rather, their solutions tend to form an integral part of the text. For this reason, most of the exercises cannot be skipped over without severely impairing the entire presentation.

The fact that this book largely ignores the many applications of quantum mechanics should not be taken to imply that these applications are irrelevant to the problem of understanding the theory. It is true that the *theory* of quantum mechanics provides a conceptual setting for the various applications, thereby interrelating these applications in a logically satisfying way; however, it is also true that the specific *applications* of quantum mechanics provide concrete examples of the highly abstract theory, thereby rendering the theory intelligible and retainable. Thus a *real* understanding of quantum mechanics can come only after both its theory *and* its applications have been thoroughly studied, each in the light of the other. Since this little book is confined to an elementary presentation of the theory only, it obviously cannot carry the reader all the way to the level of a full-fledged "quantum mechanician." However, it is hoped that the use of this book as an introduction or supplement to the more conventional textbooks and courses in quantum mechanics will help to steady and lengthen the reader's "first steps" in this journey.

CHAPTER

2

The Mathematical Language of Quantum Mechanics

Classical mechanics is formulated in terms of the mathematical language of differential and integral calculus. For example, velocity and acceleration are defined in terms of the derivative, work and impulse are defined in terms of the integral, and the conservation principles of energy and momentum find their rigorous justifications in certain elementary theorems of calculus. Quantum mechanics, too, has a mathematical language—a language that involves not only calculus but also several other branches of mathematics. In this chapter we present, in as concise and elementary a way as we can, those mathematical concepts (other than calculus) which are essential to a meaningful understanding of quantum mechanics. The necessity for achieving a reasonable degree of fluency in this mathematical language is even greater in the case of quantum mechanics than classical mechanics; for quantum theory unfortunately does not readily lend itself to nonmathematical clarifications in terms of notions familiar to us from everyday experience. The reader is therefore urged to gain a full understanding of the material presented in this chapter before proceeding to the following chapters.

2-1 PROBABILITY AND STATISTICS

In order to develop several concepts of probability theory that we shall need in our discussion of quantum mechanics, let us imagine that we have a box which contains N balls, each marked with some number which we denote generically by v. In general, the same v-number may appear on more than one ball, and we let n_k be the number of balls on which there appears the particular v-number v_k. The box of balls is therefore described by the two sets of numbers

$v_1, v_2, v_3, \ldots$, and $n_1, n_2, n_3, \ldots$. Evidently, the integers $\{n_k\}$ satisfy $\Sigma_k n_k = N$.†

Suppose we select a ball at random from the box; what is the *probability* p_k that the selected ball will show the value v_k? Since out of N possible selections, n_k of these would yield the v-number v_k, we conclude that

$$p_k = \frac{n_k}{N} \tag{2-1}$$

Thus if $n_k = 0$ it would be *impossible* to select a ball showing v_k, and we would have $p_k = 0$; on the other hand, if $n_k = N$ it would be an *absolute certainty* that the selected ball would show v_k, and we would have $p_k = 1$. In general, the numbers $\{p_k\}$ satisfy the conditions

$$0 \leqq p_k \leqq 1 \text{ for all } k \tag{2-2a}$$

and

$$\sum_k p_k = 1 \tag{2-2b}$$

Exercise 1. Prove Eqs. (2-2).

Let us calculate the probability that a single random selection from the box will yield a ball showing *either* v_k or v_j. Since out of N possible selections, a total of $(n_k + n_j)$ would yield one of these v-numbers, we conclude that

$$p \text{ (either } v_k \text{ or } v_j) = \frac{n_k + n_j}{N} = p_k + p_j \tag{2-3a}$$

In light of this result, we may view Eq. (2-2b) as simply stating that it is an absolute certainty that a randomly selected ball will show *some* v_k number.

Suppose we now make *two* random selections, taking care to return to the box the first ball selected before making the second selection (thus, it is possible to pick the same ball both times). What is the probability that the first ball will show the value v_k and the second ball then show the value v_j? There are n_k ways to select a v_k-ball, and for each of these ways there are n_j ways to select a v_j-ball; thus, there are a total of $n_k \cdot n_j$ ways to select first a v_k-ball and then a v_j-ball. However, there are N possible selections for the first ball, and for each of these there are N possible second selections; thus,

†For the sake of brevity, we shall often denote a *set* of entities $a_1, a_2, a_3, \ldots$ by $\{a_i\}$.

there are a total of $N \cdot N$ possible double selections. We conclude then that

$$p \text{ (first } v_k \text{, then } v_j) = \frac{n_k \cdot n_j}{N \cdot N} = p_k \cdot p_j \qquad (2\text{-}3b)$$

Equations (2-3) form the basis for almost all considerations involving probability theory.

Exercise 2. In the situation we have been discussing, suppose the box contains $N = 50$ balls, each bearing some integer between 1 and 8; specifically, letting n_k be the number of balls showing the value $v_k = k$, suppose that $n_1 = 3$, $n_2 = 2$, $n_3 = 5$, $n_4 = 8$, $n_5 = 13$, $n_6 = 9$, $n_7 = 6$ and $n_8 = 4$. Use the probability concepts developed above to calculate the probability that the numbers found on two random samplings will sum to 5. [*Ans.:* 68/2500]

Suppose now that we subject our box of N balls to M samplings; that is, we select a ball at random from the box, record its v-number and return it to the box a total of M times. We denote by $v^{(i)}$ the v-value recorded on the ith sampling, and we make the following two definitions: The *mean* or *average* of the v-values recorded is

$$\langle v \rangle \equiv \frac{\sum_{i=1}^{M} v^{(i)}}{M} \qquad (2\text{-}4)$$

and the *root-mean-square* (or rms) *deviation* of these values is

$$\Delta v \equiv \sqrt{\frac{\sum_{i=1}^{M} (v^{(i)} - \langle v \rangle)^2}{M}} \qquad (2\text{-}5)$$

The definition of $\langle v \rangle$ is undoubtedly familiar and needs little comment. It describes the way in which we would ordinarily compute the "best value" of a series of measurements, or the "average grade" on a class quiz. The latter analogy is actually more appropriate to our discussion here, since we evidently do *not* wish to imply that $\langle v \rangle$ has some truth or legitimacy beyond that of any of the individual $v^{(i)}$-values.

Less familiar, perhaps, than the definition of the mean value $\langle v \rangle$ is the rms deviation Δv. We see that to compute this quantity, we first calculate the deviation from the mean, $v^{(i)} - \langle v \rangle$, of each v-number obtained; we next compute the average of the squares of

these deviations (the squares are taken to keep the positive and negative deviations from canceling each other); and finally, to counteract to some extent the squaring, we take the square root of this average. Thus Δv is the square *root* of the *mean* of the *squares* of the *deviations* of the $v^{(i)}$-values from $\langle v \rangle$. This quantity might also be called the *rms dispersion*, since it evidently measures the extent to which the $v^{(i)}$-values are "dispersed" about $\langle v \rangle$. Of course, this is not the only quantity which can be calculated to measure this dispersion; for example, we could compute instead the average of the absolute values of the deviations, $|v^{(i)} - \langle v \rangle|$. However, the quantity in Eq. (2-5) has the advantage that it can be written in another often useful form. Specifically, we see from Eq. (2-5) that

$$(\Delta v)^2 = \frac{\sum_{i=1}^{M} [v^{(i)2} - 2\langle v \rangle v^{(i)} + \langle v \rangle^2]}{M}$$

$$= \frac{\sum_{i=1}^{M} v^{(i)2}}{M} - 2\langle v \rangle \frac{\sum_{i=1}^{M} v^{(i)}}{M} + \frac{M\langle v \rangle^2}{M}$$

$$= \langle v^2 \rangle - 2\langle v \rangle\langle v \rangle + \langle v \rangle^2$$

Therefore,

$$\Delta v = \sqrt{\langle v^2 \rangle - \langle v \rangle^2} \tag{2-6}$$

In words, the rms deviation of the $v^{(i)}$-values is equal to the square root of the difference between the *average of the square* and the *square of the average*. It is to be noted that these two quantities are *not* in general equal; indeed, a comparison of Eqs. (2-5) and (2-6) reveals that $\langle v^2 \rangle = \langle v \rangle^2$ *only if every* $v^{(i)}$-value coincides with $\langle v \rangle$. Equation (2-6) tells us that the extent to which $\langle v^2 \rangle$ and $\langle v \rangle^2$ differ provides us with a direct measure of the dispersion in the $v^{(i)}$-values.

If we have a knowledge of the two sets of numbers $\{v_k\}$ and $\{n_k\}$, or equivalently $\{v_k\}$ and $\{p_k\}$, it would seem that we ought to be able to *predict* approximately what values would be obtained for $\langle v \rangle$ and Δv. The key to making such a prediction is the following assumption: since n_k of the N balls have the number v_k, then in M random samplings of these balls we ought to obtain the value v_k approximately m_k times, where $m_k/M = n_k/N$. Thus, using Eq. (2-1), the approximate number of times the value v_k should appear in the set of values $v^{(1)}, v^{(2)}, \ldots, v^{(M)}$ is

$$m_k = \frac{n_k}{N} M = p_k M$$

With this, the sum in Eq. (2-4) can be written

$$\sum_{i=1}^{M} v^{(i)} = \sum_k m_k v_k = \sum_k (p_k M) v_k$$

and Eq. (2-4) becomes

$$\langle v \rangle = \sum_k p_k v_k \tag{2-7}$$

Equation (2-7) expresses $\langle v \rangle$ as a "weighted sum" of the possible v_k-values; the weight assigned to any particular value v_k is just the probability of its occurrence, p_k. It should be remarked that this value for $\langle v \rangle$ is the "theoretically expected" value; the "experimental" value in Eq. (2-4) will generally differ somewhat from this theoretical value owing to the randomness involved. However, in the limit of very many experimental samplings ($M \longrightarrow \infty$), the value in Eq. (2-4) may be expected to get arbitrarily close to the value in Eq. (2-7).

Equation (2-7) may be generalized quite easily, as the following exercise shows.

Exercise 3. Let f be a given function of v, and let this function be evaluated for each of the $v^{(i)}$-values. Prove that the average or mean of the resulting set of $f(v^{(i)})$-values is

$$\langle f(v) \rangle = \sum_k p_k \, f(v_k) \tag{2-8}$$

[Note that by putting $f(v) = v$ in Eq. (2-8), we obtain Eq. (2-7).]

By putting $f(v) = v^2$ in Eq. (2-8), we see that

$$\langle v^2 \rangle = \sum_k p_k {v_k}^2$$

Using this and Eq. (2-7), we may thus write Eq. (2-6) as

$$\Delta v = \sqrt{\left(\sum_k p_k {v_k}^2\right) - \left(\sum_k p_k v_k\right)^2} \tag{2-9}$$

We now observe that Eqs. (2-7) and (2-9) express the two basic quantities $\langle v \rangle$ and Δv wholly in terms of the numbers $\{v_k\}$ and $\{p_k\}$. Thus, given a set of values $v_1, v_2, \ldots$ distributed with probabilities $p_1, p_2, \ldots$, Eqs. (2-7) and (2-9) allow us to calculate the theoretically expected *mean* and *rms deviation* to be obtained in any random sampling of these v-values.

Exercise 4. Consider the collection of numbered balls described in Exercise 2.

(a) Calculate $\langle v \rangle$ and Δv. [*Ans.:* $\langle v \rangle = 4.94$ and $\Delta v = 1.8$]

(b) Sketch a "frequency bar-graph" of the expected results of $M = 100$ samplings [i.e., lay out the values v_k on the horizontal axis, and construct vertical "bars" to indicate the number of times each v_k-value should be obtained.] Show on the graph by means of a vertical line the value $\langle v \rangle$. Also, draw a horizontal line of length $2\Delta v$ in such a way that it indicates roughly the "spread" or "dispersion" of the v-values about $\langle v \rangle$.

The foregoing exercise illustrates the overall significance of $\langle v \rangle$ and Δv. Certainly a *complete* description of the expected results of a "multiple sampling" experiment requires the specification of *all* the numbers (v_1, p_1), (v_2, p_2), (v_3, p_3), ... However, if we are asked to describe the results with *only two* numbers, we would evidently do well to state the values of $\langle v \rangle$ and Δv: $\langle v \rangle$ is essentially a "collective value" for the set of v-numbers, while Δv (or the smallness thereof) provides a quantitative measure of the degree to which it is actually meaningful to so characterize the *set* of v-values by a *single* value. These ideas will play a very important role in understanding certain basic concepts in quantum mechanics.

2-2 COMPLEX NUMBERS

An understanding of quantum mechanics requires some knowledge of a few elementary properties of complex numbers. For our purposes, we may define a complex number c as a quantity which can be written

$$c \equiv a + ib \tag{2-10a}$$

Here a and b are ordinary real numbers, while the "number" i satisfies

$$i^2 \equiv -1 \quad \text{or} \quad i \equiv \sqrt{-1} \tag{2-10b}$$

The real number a is called the "real part" of c, and the real number b is called the "imaginary part" of c:

$$a \equiv \mathrm{Re}c \qquad b \equiv \mathrm{Im}c \tag{2-10c}$$

If $b = 0$, then c is said to be a "pure real" number; if $a = 0$, then c is said to be a "pure imaginary" number. We write $c = 0$ if and only if $a = b = 0$.

Complex numbers can be added and multiplied. The rules for carrying out these operations are the same as for ordinary real numbers, but taking account of Eq. (2-10b). Thus if $c_1 = a_1 + ib_1$ and $c_2 = a_2 + ib_2$, then the *sum* of c_1 and c_2 is defined to be the complex number

$$c_1 + c_2 \equiv (a_1 + ib_1) + (a_2 + ib_2) = (a_1 + a_2) + i(b_1 + b_2) \qquad (2\text{-}11a)$$

and the *product* of c_1 times c_2 is defined to be the complex number

$$c_1 \cdot c_2 \equiv (a_1 + ib_1) \cdot (a_2 + ib_2) = (a_1\, a_2 - b_1\, b_2) + i(a_1\, b_2 + b_1\, a_2) \qquad (2\text{-}11b)$$

If c is written in the form of Eq. (2-10a), then the *complex conjugate* of c is defined to be the complex number

$$c^* \equiv a - ib \qquad (2\text{-}12)$$

Exercise 5. Prove the following properties of the complex conjugate:

(a) $\mathrm{Re}c = \dfrac{c + c^*}{2} \qquad \mathrm{Im}c = \dfrac{c - c^*}{2i}$ (2-13)

(b) c is a pure *real* number if and only if $c^* = c$. c is a pure *imaginary* number if and only if $c^* = -c$.

(c) $c^{**} = c$ (2-14a)

$(c_1 + c_2)^* = c_1^* + c_2^*$ (2-14b)

$(c_1 c_2)^* = c_1^* c_2^*$ (2-14c)

The *square modulus* of c is denoted by $|c|^2$ and is defined to be the product of c times its complex conjugate:

$$|c|^2 \equiv c^* c \qquad (2\text{-}15a)$$

The *modulus* of a complex number is just the positive square root of its square modulus:

$$|c| \equiv +\sqrt{c^* c} \qquad (2\text{-}15b)$$

Exercise 6. Prove the following properties of the modulus:

(a) $|c|^2 = (\mathrm{Re}c)^2 + (\mathrm{Im}c)^2$ or $|c| = \sqrt{(\mathrm{Re}c)^2 + (\mathrm{Im}c)^2}$ (2-16)

(b) $|c| \geqq |\mathrm{Re}c|$ and $|c| \geqq |\mathrm{Im}c|$ (2-17)

(c) $|c_1 c_2| = |c_1|\,|c_2|$ (2-18a)

$|c_1 + c_2| \leqq |c_1| + |c_2|$ (2-18b)

[*Hint:* Write $c_1 = a_1 + ib_1$ and $c_2 = a_2 + ib_2$ for part (c).]

The *square modulus* of c should not be confused with the *square* of c. If $c = a + ib$, then using Eq. (2-11b), we have

$$|c|^2 = c^*c = a^2 + b^2$$

whereas

$$c^2 = cc = (a^2 - b^2) + i(2ab)$$

Clearly, *the square modulus is always a nonnegative real number*, while the square is in general complex. In fact, the distinction between the square and the square modulus disappears if and only if c is pure real; in this case, the modulus becomes the absolute value.

It should be remarked that the complex number system can be set up without making use of the "number" $\sqrt{-1}$. This is done by initially *defining* a complex number c to be an ordered pair of real numbers (a,b), and then setting forth appropriate rules for algebraically manipulating these ordered pairs. For example, if $c_1 = (a_1, b_1)$ and $c_2 = (a_2, b_2)$, then the sum $c_1 + c_2$ would be *defined* as the ordered pair $(a_1 + a_2, b_1 + b_2)$, and the product $c_1 \cdot c_2$ would be *defined* as the ordered pair $(a_1 a_2 - b_1 b_2, a_1 b_2 + a_2 b_1)$. The symbol i can then be introduced as merely an *ad hoc* device to simplify these rules; that is, by writing the ordered pair (a,b) as $a + ib$, we can manipulate complex numbers using the familiar rules of the algebra of *real* numbers, with the additional rule that i^2 is always to be replaced by -1. Although we shall always write $a + ib$ instead of (a,b), the reader should try to adopt this essentially "algebraic" attitude toward complex numbers instead of the popular "mystical" attitude, which is overly concerned with the square roots of negative real numbers.

In exact analogy with the foregoing, we can define a *complex function* ψ of a *real variable* x to be a function of the form

$$\psi(x) = u(x) + iv(x) \tag{2-19}$$

where $u(x)$ and $v(x)$ are ordinary real functions of the real variable x. All the preceding equations concerning the complex number c hold for the complex function $\psi(x)$, provided that we replace Rec by Re$\psi(x) = u(x)$, and Imc by Im$\psi(x) = v(x)$. Thus, for example, the complex conjugate of $\psi(x)$ is $\psi^*(x) = u(x) - iv(x)$, and the square modulus of $\psi(x)$ is $|\psi(x)|^2 = u^2(x) + v^2(x)$.

The complex function $\psi(x)$ can be differentiated and integrated with respect to its argument x. The rules for carrying out these two operations are just what one would expect:

$$\frac{d}{dx}\psi(x) \equiv \frac{d}{dx}u(x) + i\frac{d}{dx}v(x)$$

and

$$\int_a^b \psi(x)dx \equiv \int_a^b u(x)dx + i\int_a^b v(x)dx$$

We note that the derivative of $\psi(x)$ is a complex function of x, whereas the definite integral of $\psi(x)$ is a complex number, in exact analogy with the situation for real functions of x.

It should perhaps be mentioned that it is also possible to define complex functions ψ of a *complex variable* $z = x + iy$. The situation with respect to differentiation and integration then becomes rather involved. However, in this book we shall require only a knowledge of the comparatively simple properties of complex functions of a real variable, as outlined above.

Exercise 7. The complex function e^{ikx} (k real) is *defined* by

$$e^{ikx} \equiv \cos kx + i \sin kx \tag{2-20a}$$

Prove from this definition that e^{ikx} has the following properties:

$$(e^{ikx})^* = e^{-ikx} \tag{2-20b}$$

$$e^{ik_1x} \cdot e^{ik_2x} = e^{i(k_1+k_2)x} \tag{2-20c}$$

$$|e^{ikx}|^2 = 1 \tag{2-20d}$$

$$\frac{d}{dx} e^{ikx} = (ik)\, e^{ikx} \tag{2-20e}$$

$$\int e^{ikx}\, dx = \frac{1}{ik} e^{ikx} + C \tag{2-20f}$$

[We sometimes will write $\exp(ikx)$ instead of e^{ikx}.]

2-3 HILBERT SPACE VECTORS

The language of quantum mechanics is mainly the language of a branch of mathematics called "vector spaces." The reader is assumed to be familiar with the elementary properties of "ordinary vectors" in three-dimensional Euclidean space (mathematicians denote this space symbolically by $\mathcal{E}_3$). Actually, the notion of a vector space is much more general than this. In fact, quantum mechanics is formulated in terms of an *infinite*-dimensional vector space called a "Hilbert space" (denoted symbolically by $\mathcal{H}$); in $\mathcal{H}$ the "vectors" are *not* directed line segments, as in $\mathcal{E}_3$, but rather are *complex functions of*

real variables. A complete development of the mathematics of the Hilbert space is beyond our reach here; however, at the expense of a little mathematical rigor and generality, we shall find it possible to come to a fairly good understanding of the Hilbert space by drawing suitable analogies with the simpler, more familiar properties of $\mathfrak{E}_3$. To this end, we shall begin our discussion of vectors in $\mathcal{H}$ by reviewing some of the salient ideas concerning vectors in $\mathfrak{E}_3$.

A "vector" in $\mathfrak{E}_3$ can be defined as a *directed line segment.* Thus, a vector $\mathbf{v}$ in $\mathfrak{E}_3$ possesses the properties of magnitude and direction; the magnitude of $\mathbf{v}$, written $|\mathbf{v}|$, is the length of the line segment, and the direction of $\mathbf{v}$ is the direction of travel from the "tail" of the line segment to the "head" of the line segment.

Two operations common to all vector spaces are the operations of *scalar multiplication* and *vector addition. Scalars* in $\mathfrak{E}_3$ are simply the set of all *real numbers.* The multiplication of a vector $\mathbf{v}$ by a scalar r yields a new vector, written $r\mathbf{v}$, whose direction is the same as that of $\mathbf{v}$ but whose magnitude is $|r|$ times the magnitude of $\mathbf{v}$ ($|r\mathbf{v}| = |r|\,|\mathbf{v}|$); negative scalar multipliers reverse the direction. The addition of two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ yields a new vector, which is written $\mathbf{v}_1 + \mathbf{v}_2$; this vector is obtained by placing the tail of $\mathbf{v}_2$ at the head of $\mathbf{v}_1$, and then constructing the directed line segment from the tail of $\mathbf{v}_1$ to the head of $\mathbf{v}_2$. These two operations of scalar multiplication and vector addition allow us to form *linear combinations* of vectors; thus, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are any two vectors in $\mathfrak{E}_3$, and r_1 and r_2 are any two $\mathfrak{E}_3$-scalars (i.e., real numbers), then the "linear combination"

$$\mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 \tag{2-21}$$

is a well-defined vector in $\mathfrak{E}_3$.

Another important feature of many (but not all) vector spaces is the existence of an operation called the *inner product.* In $\mathfrak{E}_3$ the inner product of two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is written $\mathbf{v}_1 \cdot \mathbf{v}_2$, and is customarily called the "dot product" of $\mathbf{v}_1$ and $\mathbf{v}_2$. By *definition*,

$$\mathbf{v}_1 \cdot \mathbf{v}_2 \equiv |\mathbf{v}_1|\,|\mathbf{v}_2| \cos\theta_{12} \tag{2-22}$$

where θ_{12} is the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$ when these two vectors are placed tail-to-tail. Geometrically, $\mathbf{v}_1 \cdot \mathbf{v}_2$, can be thought of as the product of the length of $\mathbf{v}_1$ times the projected length $|\mathbf{v}_2| \cos\theta_{12}$ of $\mathbf{v}_2$ on $\mathbf{v}_1$, or equivalently as the product of the length of $\mathbf{v}_2$ times the projected length $|\mathbf{v}_1| \cos\theta_{12}$ of $\mathbf{v}_1$ on $\mathbf{v}_2$.

It is evident from Eq. (2-22) that *the inner product of two vectors is always a scalar* (in this case, a real number). In particular, the inner product of a vector with itself, called the *norm* of the

vector, is always a *nonnegative real number*:

$$\text{Norm of } \mathbf{v} \equiv \mathbf{v} \cdot \mathbf{v} = |\mathbf{v}|^2 \geqq 0 \tag{2-23}$$

It can be shown from Eq. (2-22) that the inner product in $\mathfrak{E}_3$ satisfies the following relations:

$$\mathbf{v}_2 \cdot \mathbf{v}_1 = \mathbf{v}_1 \cdot \mathbf{v}_2 \tag{2-24a}$$

$$r_1 \mathbf{v}_1 \cdot r_2 \mathbf{v}_2 = r_1 r_2 \mathbf{v}_1 \cdot \mathbf{v}_2 \tag{2-24b}$$

$$(\mathbf{v}_1 + \mathbf{v}_2) \cdot (\mathbf{v}_3 + \mathbf{v}_4) = \mathbf{v}_1 \cdot \mathbf{v}_3 + \mathbf{v}_1 \cdot \mathbf{v}_4 + \mathbf{v}_2 \cdot \mathbf{v}_3 + \mathbf{v}_2 \cdot \mathbf{v}_4 \tag{2-24c}$$

$$|\mathbf{v}_1 \cdot \mathbf{v}_2| \leqq \sqrt{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\sqrt{\mathbf{v}_2 \cdot \mathbf{v}_2} \tag{2-24d}$$

Exercise 8. Verify the foregoing relations.

Equation (2-24a) says that the inner product is cummutative. Equations (2-24b) and (2-24c) show how the inner product behaves with respect to the operations of scalar multiplication and vector addition. Equation (2-24d) states a very fundamental property of the inner product; this relation is often referred to as the Schwarz inequality.

Two vectors are said to be *orthogonal* if they are perpendicular to each other. Since $\cos(\pi/2) = 0$, then from Eq. (2-22) we can write

$$\mathbf{v}_1 \text{ and } \mathbf{v}_2 \text{ are } \textit{orthogonal} \text{ if and only if } \quad \mathbf{v}_1 \cdot \mathbf{v}_2 = 0 \tag{2-25}$$

Indeed, we may adopt this as our *definition* of orthogonality, if we agree to regard the null vector **0** as being orthogonal to any other vector and also to itself.

We often deal with *sets* of vectors, $\mathbf{v}_1, \mathbf{v}_2, \ldots$ or more compactly, $\{\mathbf{v}_i\}$. Concerning such sets we make the following definitions:

(i) The set $\{\mathbf{v}_i\}$ is said to be *orthonormal* if and only if each vector of the set is orthogonal to every other vector of the set, and each vector of the set has unit norm. These properties can be expressed most succinctly through the use of the so-called "Kronecker delta symbol," δ_{ij}, which is defined by

$$\delta_{ij} \equiv \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \tag{2-26}$$

Thus we have

$$\{\mathbf{v}_i\} \text{ is an } \textit{orthonormal set} \text{ if and only if } \quad \mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij} \tag{2-27}$$

(ii) The set $\{\mathbf{v}_i\}$ is said to be *complete* if and only if any vector in $\mathfrak{E}_3$ can be written as a linear combination of the vectors in $\{\mathbf{v}_i\}$. In other words, $\{\mathbf{v}_i\}$ is a complete set if and only if, for any vector **v** in

$\mathscr{E}_3$, there exists at least one set of scalars $\{r_i\}$ such that $\mathbf{v} = \Sigma_i r_i \mathbf{v}_i$. It turns out that, in $\mathscr{E}_3$, any set of three or more noncoplanar vectors constitutes a complete set.

Of particular interest are those sets of vectors which are *both* orthonormal and complete; such a set is called an *orthonormal basis set*. In $\mathscr{E}_3$, there are infinitely many different orthonormal basis sets (they differ from one another by simple rotations), and all such sets have exactly *three* vectors; for this reason, $\mathscr{E}_3$ is said to be "three-dimensional." A specific orthonormal basis set in $\mathscr{E}_3$ is usually written as $(\mathbf{x},\mathbf{y},\mathbf{z})$ or $(\mathbf{i},\mathbf{j},\mathbf{k})$ or $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$; we shall use the latter notation. Thus, we have from the orthonormality of the set $\{\mathbf{e}_i\}$,

$$\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} \qquad (i,j = 1,2,3) \tag{2-28}$$

Moreover, since $\{\mathbf{e}_i\}$ is complete, then given any vector $\mathbf{v}$ we can find scalars r_1, r_2 and r_3 such that

$$\mathbf{v} = \sum_{i=1}^{3} r_i \mathbf{e}_i \tag{2-29a}$$

Indeed, using Eqs. (2-24) and (2-28), we see that

$$\mathbf{e}_j \cdot \mathbf{v} = \mathbf{e}_j \cdot \left(\sum_{i=1}^{3} r_i \mathbf{e}_i \right) = \sum_{i=1}^{3} r_i (\mathbf{e}_j \cdot \mathbf{e}_i) = \sum_{i=1}^{3} r_i \delta_{ij} = r_j \tag{2-29b}$$

That is, the *expansion coefficients* or *components* r_i of $\mathbf{v}$ in the orthonormal basis $\{\mathbf{e}_i\}$ are just the scalars $\mathbf{e}_i \cdot \mathbf{v}$. Therefore, we can write Eq. (2-29a) as

$$\mathbf{v} = \sum_{i=1}^{3} (\mathbf{e}_i \cdot \mathbf{v}) \mathbf{e}_i \tag{2-29c}$$

an equation which may be regarded as an identity for all $\mathbf{v}$ in $\mathscr{E}_3$, and all orthonormal basis sets $\{\mathbf{e}_i\}$. The import of Eq. (2-29c) is illustrated in Fig. 1.

Exercise 9. If two $\mathscr{E}_3$-vectors $\mathbf{a}$ and $\mathbf{b}$ have components $\{a_i\}$ and $\{b_i\}$ relative to a given basis $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$, prove that

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{3} a_i b_i \tag{2-30a}$$

and in particular that

$$\mathbf{a} \cdot \mathbf{a} = \sum_{i=1}^{3} a_i^2 \tag{2-30b}$$

[*Hint:* Make use of Eqs. (2-24) and (2-28).]

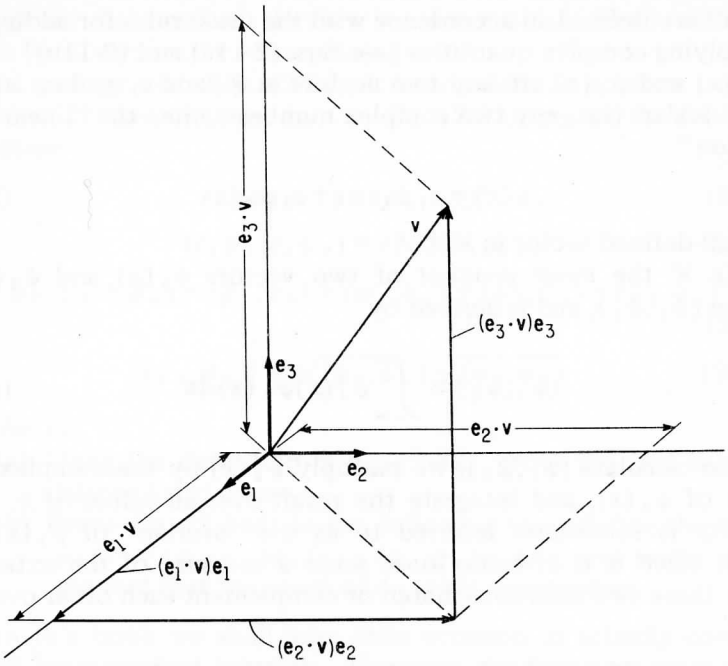

Fig. 1. Illustrating the expansion of an arbitrary vector v in an orthonormal basis set e_1, e_2, e_3. The inner product $(e_i \cdot v)$ is just the projected length or "component" of v in the direction of e_i; therefore, v can be written as the vector sum $v = \sum_{i=1}^{3} (e_i \cdot v)e_i$.

The preceding discussion of vectors in $\mathcal{E}_3$ is by no means complete, but it is extensive enough for our purposes. We shall now show how these vector concepts in $\mathcal{E}_3$ carry over into the Hilbert space $\mathcal{H}$.

We *define* a vector in $\mathcal{H}$ to be a *complex function* ψ *of a single real variable* x [see Eq. (2-19)]. In other words, a vector in $\mathcal{H}$ is a *rule of correspondence* which assigns to each real number x a complex number $\psi(x)$. To be precise, we should mention that not all such functions are truly vectors in $\mathcal{H}$, but only those functions that satisfy a certain condition; we shall state and discuss this condition later [see Eq. (2-35)].

The *scalars* in $\mathcal{H}$ are by definition the set of all *complex numbers.* This is to be contrasted with the situation in $\mathcal{E}_3$, where the scalars are the set of all *real* numbers.

The two operations of "scalar multiplication" and "vector ad-

dition" are defined in accordance with the usual rules for adding and multiplying complex quantities [see Eqs. (2-11a) and (2-11b)]. Thus, if $\psi_1(x)$ and $\psi_2(x)$ are any two vectors in $\mathcal{H}$, and c_1 and c_2 are any two $\mathcal{H}$-scalars (i.e., any two complex numbers), then the "linear combination"

$$\psi(x) = c_1\psi_1(x) + c_2\psi_2(x) \tag{2-31}$$

is a well-defined vector in $\mathcal{H}$.†

In $\mathcal{H}$ the *inner product* of two vectors $\psi_1(x)$ and $\psi_2(x)$ is written (ψ_1,ψ_2), and is *defined* by

$$(\psi_1,\psi_2) \equiv \int_{-\infty}^{\infty} \psi_1^*(x)\psi_2(x)\,dx \tag{2-32}$$

Thus to calculate (ψ_1,ψ_2), we multiply $\psi_2(x)$ by the complex conjugate of $\psi_1(x)$, and integrate the result over all values of x. This quantity is sometimes referred to as the "overlap" of $\psi_1(x)$ and $\psi_2(x)$, since it is in some loose sense a measure of the extent to which these two functions match or complement each other over the x-axis.

Exercise 10. Write $\psi_j(x) = u_j(x) + iv_j(x)$, for $j = 1$ and $j = 2$, and show explicitly that (ψ_1,ψ_2) is an $\mathcal{H}$-scalar. Note in particular that if x had *not* been integrated over in Eq. (2-32) then (ψ_1,ψ_2) would *not* have been an $\mathcal{H}$-scalar.

The fact that (ψ_1,ψ_2) is always an $\mathcal{H}$-scalar parallels the fact that $\mathbf{v}_1 \cdot \mathbf{v}_2$ is always an $\mathcal{E}_3$-scalar.

In $\mathcal{E}_3$ we defined the norm of a vector to be its inner product with itself. Analogously, in $\mathcal{H}$ we define the *norm* of the vector $\psi(x)$ to be

$$\text{Norm of } \psi(x) \equiv (\psi,\psi) = \int_{-\infty}^{\infty} \psi^*(x)\psi(x)\,dx = \int_{-\infty}^{\infty} |\psi(x)|^2\,dx \geq 0 \tag{2-33}$$

Exercise 11. Write $\psi(x) = u(x) + iv(x)$, and show explicitly that (ψ,ψ) is a nonnegative real number. Note in particular that if one of the functions in the integrand in Eq. (2-32) had *not* been complex conjugated, then (ψ,ψ) would *not* always be a real number.

†It should be emphasized that, in denoting a Hilbert space vector by the symbol "$\psi(x)$," we are referring to the *function* ψ of the variable x, and *not* the *value* of this function at the point x.

We see then that in both $\mathcal{E}_3$ and $\mathcal{H}$, the norm of a vector is a nonnegative real number. In fact, the norm of a vector is zero only for the "null vectors," $\mathbf{v} = \mathbf{0}$ and $\psi(x) \equiv 0$.

In analogy with Eqs. (2-24), it is not difficult to show that the definition of the inner product in Eq. (2-32) implies the following properties:

$$(\psi_2, \psi_1) = (\psi_1, \psi_2)^* \tag{2-34a}$$

$$(c_1 \psi_1, c_2 \psi_2) = c_1^* c_2 (\psi_1, \psi_2) \tag{2-34b}$$

$$(\psi_1 + \psi_2, \psi_3 + \psi_4) = (\psi_1, \psi_3) + (\psi_1, \psi_4) + (\psi_2, \psi_3) + (\psi_2, \psi_4) \tag{2-34c}$$

$$|(\psi_1, \psi_2)| \leqq \sqrt{(\psi_1, \psi_1)}\sqrt{(\psi_2, \psi_2)} \tag{2-34d}$$

Exercise 12.

(a) Using the definition in Eq. (2-32), prove the first three relations listed above.

(b) Prove Eq. (2-34d) in the following way: define the vector $\psi_3(x) \equiv (\psi_1, \psi_2)\psi_1(x) - (\psi_1, \psi_1)\psi_2(x)$, and make use of the fact that the norm of $\psi_3(x)$ is nonnegative.

In this book we shall have little occasion to actually *compute* explicit inner-product integrals. However, the four *properties* of the inner product integral listed in Eqs. (2-34) will be used quite extensively, so the reader should become as familiar with them as possible. We note in particular the appearance of the complex conjugation operation in Eqs. (2-34a) and (2-34b), in contrast to Eqs. (2-24a) and (2-24b); however, even in the latter equations, the complex conjugation operation would evidently not be incorrect, but merely unnecessary. The Schwarz inequality in Eq. (2-34d) has the same form as in Eq. (2-24d), but it should be noted that $|\mathbf{v}_1 \cdot \mathbf{v}_2|$ in Eq. (2-24d) means the *absolute value* of the (possibly negative) *real* number $\mathbf{v}_1 \cdot \mathbf{v}_2$, whereas $|(\psi_1, \psi_2)|$ in Eq. (2-34d) means the *modulus* of the (in general) *complex* number (ψ_1, ψ_2).

The reader should now begin to appreciate the rationale for calling complex functions "vectors" in a vector space. From a *strictly mathematical* point of view, directed line segments may be regarded as "vectors," not because they possess the properties of magnitude and direction, but rather because we can define for directed line segments the three operations of scalar multiplication, vector addition, and vector inner multiplication, in such a way that Eqs. (2-23) and (2-24) are obeyed. It is these latter relations that actually determine the "vector character" of directed line segments, and not the concept of a directed line segment itself, nor even the

specific recipes for forming the scalar product, vector sum and vector inner product, relative to directed line segments. Now we have just seen that, if we adopt certain well-defined rules for obtaining the "scalar product," "vector sum" and "inner product," relative to *complex functions of a real variable*, then we arrive at the properties expressed in Eqs. (2-33) and (2-34). These properties are essentially identical to those in Eqs. (2-23) and (2-24); consequently, we are entirely justified in regarding complex functions as "vectors" in a vector space. In particular, our definition of the inner product in Eq. (2-32), which at first sight probably seemed rather peculiar to the reader, was chosen simply because it was a way of obtaining a unique scalar from two vectors such that Eqs. (2-33) and (2-34) were satisfied. If we could conjure up a *different* set of rules for forming "linear combinations" and "inner products" of complex functions, which still satisfied all the conditions in Eqs. (2-33) and (2-34), then we would have thereby constructed *another* perfectly valid "vector space" of complex functions; however, that vector space would probably not turn out to be as relevant to the task of describing physical phenomena as our present "Hilbert space" turns out to be.

We are now in a position to state the condition alluded to earlier which a function $\psi(x)$ must satisfy in order to be a vector in $\mathcal{H}$. We admit as vectors of $\mathcal{H}$ *only* those functions $\psi(x)$ which have a *finite norm*:

$$\psi(x) \text{ is a vector of } \mathcal{H} \text{ if and only if } (\psi,\psi) = \int_{-\infty}^{\infty} |\psi(x)|^2\, dx < \infty \tag{2-35}$$

We should note that an analogous condition was implicitly imposed on $\mathcal{E}_3$-vectors, through their definition as directed line *segments* (i.e., lines of finite length). Condition (2-35) insures the following two important results:

(i) If $\psi_1(x)$ and $\psi_2(x)$ are any two vectors in $\mathcal{H}$, then the inner product (ψ_1,ψ_2) "exists" in the sense that it is a complex number with a finite modulus. To see that this is true, we note that, since $\psi_1(x)$ and $\psi_2(x)$ are $\mathcal{H}$-vectors, then $(\psi_1,\psi_1) < \infty$ and $(\psi_2,\psi_2) < \infty$ by Eq. (2-35). Our result then follows immediately from the Schwarz inequality:

$$|(\psi_1,\psi_2)| \leqq \sqrt{(\psi_1,\psi_1)}\,\sqrt{(\psi_2,\psi_2)} < \infty$$

(ii) If $\psi_1(x)$ and $\psi_2(x)$ are any two vectors in $\mathcal{H}$, then so is any linear combination $\psi(x) = c_1\psi_1(x) + c_2\psi_2(x)$. To see that this is true, we write

$$\begin{aligned}(\psi,\psi) &= (c_1\psi_1 + c_2\psi_2,\, c_1\psi_1 + c_2\psi_2) \\ &= c_1^* c_1 (\psi_1,\psi_1) + c_1^* c_2 (\psi_1,\psi_2) + c_2^* c_1 (\psi_2,\psi_1) + c_2^* c_2 (\psi_2,\psi_2)\end{aligned}$$

$$= |c_1|^2(\psi_1,\psi_1) + |c_2|^2(\psi_2,\psi_2) + c_1^* c_2(\psi_1,\psi_2) + [c_1^* c_2(\psi_1,\psi_2)]^*$$

$$= |c_1|^2(\psi_1,\psi_1) + |c_2|^2(\psi_2,\psi_2) + 2\mathrm{Re}[c_1^* c_2(\psi_1,\psi_2)]$$

so

$$(\psi,\psi) \leqq |c_1|^2(\psi_1,\psi_1) + |c_2|^2(\psi_2,\psi_2) + 2|c_1^* c_2||(\psi_1,\psi_2)|$$

Now the first two terms on the right are bounded by virtue of Eq. (2-35), and the third term is bounded by virtue of result (i) above. Consequently, $(\psi,\psi) < \infty$, and so by Eq. (2-35) the linear combination $\psi(x) = c_1\psi_1(x) + c_2\psi_2(x)$ is indeed a vector of $\mathcal{H}$.

The preceding two results mean, first, that we can be assured that the improper integrals appearing in the definition of the inner product always converge (or make sense) for Hilbert space functions, and second, that the operation of taking linear combinations of Hilbert space functions cannot produce a non-Hilbert space function.

In direct analogy with Eq. (2-25), two $\mathcal{H}$-vectors are said to be *orthogonal* if their inner product vanishes:

$$\psi_1(x) \text{ and } \psi_2(x) \text{ are } \textit{orthogonal} \text{ if and only if } (\psi_1, \psi_2) = 0 \tag{2-36}$$

A *set* of $\mathcal{H}$-vectors $\{\psi_i(x)\}$ is said to be an *orthonormal set* if and only if each vector of the set is orthogonal to every other vector of the set, and each vector of the set has unit norm. Using the Kronecker delta symbol defined in Eq. (2-26), we therefore have in analogy to Eq. (2-27),

$$\{\psi_i(x)\} \text{ is an } \textit{orthonormal set} \text{ if and only if } (\psi_i,\psi_j) = \delta_{ij} \tag{2-37}$$

A set of $\mathcal{H}$-vectors $\{\psi_i(x)\}$ is said to be a *complete set* if and only if any vector in $\mathcal{H}$ can be written as a linear combination of the vectors in $\{\psi_i(x)\}$. In other words, $\{\psi_i(x)\}$ is a complete set if and only if, for any vector $\psi(x)$ in $\mathcal{H}$, there exists at least one set of scalars $\{c_i\}$ such that $\psi(x) = \Sigma_i c_i\psi_i(x)$.

Special use will be made of sets of $\mathcal{H}$-vectors which are *both* orthonormal and complete; such a set is called an *orthonormal basis set*. In $\mathcal{H}$, as in $\mathcal{E}_3$, there are infinitely many such orthonormal basis sets. However, whereas in $\mathcal{E}_3$ all such sets contain exactly three vectors, it turns out that in $\mathcal{H}$ all orthonormal basis sets contain infinitely many vectors; for this reason $\mathcal{H}$ is said to be *infinite-dimensional*. If $\{\epsilon_i(x)\}$ is an orthonormal basis set, then we have, by the orthonormality of the set,

$$(\epsilon_i,\epsilon_j) = \delta_{ij} \qquad (i,j = 1,2,\ldots) \tag{2-38}$$

Moreover, since $\{\epsilon_i(x)\}$ is a complete set, then given any $\mathcal{H}$-vector $\psi(x)$, we can find a set of scalars $\{c_i\}$ such that

$$\psi(x) = \sum_{i=1}^{\infty} c_i \epsilon_i(x) \tag{2-39a}$$

Indeed, using Eqs. (2-34) and (2-38), we see that

$$(\epsilon_j, \psi) = \left(\epsilon_j, \sum_{i=1}^{\infty} c_i \epsilon_i\right) = \sum_{i=1}^{\infty} c_i (\epsilon_j, \epsilon_i) = \sum_{i=1}^{\infty} c_i \delta_{ij} = c_j \tag{2-39b}$$

That is, the *expansion coefficients* or *components* c_i of $\psi(x)$ in the orthonormal basis $\{\epsilon_i(x)\}$ are just the scalars (ϵ_i, ψ). Therefore, we can write Eq. (2-39a) as

$$\psi(x) = \sum_{i=1}^{\infty} (\epsilon_i, \psi) \epsilon_i(x) \tag{2-39c}$$

an equation which may be regarded as an identity for all $\psi(x)$ in $\mathcal{H}$, and all orthonormal basis sets $\{\epsilon_i(x)\}$.

Exercise 13. If two $\mathcal{H}$-vectors $\psi(x)$ and $\phi(x)$ have components $\{c_i\}$ and $\{d_i\}$ relative to a given basis $\{\epsilon_i(x)\}$, prove that

$$(\psi, \phi) = \sum_{i=1}^{\infty} c_i^* d_i \tag{2-40a}$$

and in particular that

$$(\psi, \psi) = \sum_{i=1}^{\infty} |c_i|^2 \tag{2-40b}$$

Compare these results with those of Exercise 9.

The properties of the vector space $\mathcal{H}$ are in many respects deeper and subtler than the foregoing development would seem to indicate. However, the depth and rigor of our presentation here will be sufficient for the purposes of this book. The main results of this section are reviewed and summarized in the accompanying table. In the remainder of this book, we shall be concerned only with vectors in $\mathcal{H}$, and not vectors in $\mathcal{E}_3$. However, the correspondences which we have traced between the two vector spaces will often allow us to "visualize," by analogy with $\mathcal{E}_3$, just what it is that we are doing in $\mathcal{H}$. This will help us to keep our feet on the ground, so to speak, as we proceed through the rather abstract theory of quantum mechanics.

Correspondences Between Vectors in $\mathcal{E}_3$ and Vectors in $\mathcal{H}$

Item	$\mathcal{E}_3$	$\mathcal{H}$
Vector	Directed line segment, $\mathbf{v}$	Complex function, $\psi(x)$
Scalar	Real number, r	Complex number, c
Linear combination	$\mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2$	$\psi(x) = c_1\psi_1(x) + c_2\psi_2(x)$
Inner product[1]	$\mathbf{v}_1 \cdot \mathbf{v}_2 \equiv \lvert\mathbf{v}_1\rvert\,\lvert\mathbf{v}_2\rvert \cos\theta_{12}$	$(\psi_1,\psi_2) \equiv \int_{-\infty}^{\infty} \psi_1^*(x)\psi_2(x)dx$
Norm[2]	$\mathbf{v} \cdot \mathbf{v} = \lvert\mathbf{v}\rvert^2$	$(\psi,\psi) = \int_{-\infty}^{\infty} \lvert\psi(x)\rvert^2\,dx$
Properties of the inner product	$\mathbf{v}_2 \cdot \mathbf{v}_1 = \mathbf{v}_1 \cdot \mathbf{v}_2$ $r_1\mathbf{v}_1 \cdot r_2\mathbf{v}_2 = r_1r_2\mathbf{v}_1 \cdot \mathbf{v}_2$ $(\mathbf{v}_1 + \mathbf{v}_2) \cdot (\mathbf{v}_3 + \mathbf{v}_4)$ $= \mathbf{v}_1 \cdot \mathbf{v}_3 + \mathbf{v}_1 \cdot \mathbf{v}_4 + \mathbf{v}_2 \cdot \mathbf{v}_3 + \mathbf{v}_2 \cdot \mathbf{v}_4$ $\lvert\mathbf{v}_1 \cdot \mathbf{v}_2\rvert \leqq \sqrt{\mathbf{v}_1 \cdot \mathbf{v}_1}\sqrt{\mathbf{v}_2 \cdot \mathbf{v}_2}$	$(\psi_2,\psi_1) = (\psi_1,\psi_2)^*$ $(c_1\psi_1, c_2\psi_2) = c_1^*c_2(\psi_1,\psi_2)$ $(\psi_1 + \psi_2, \psi_3 + \psi_4)$ $= (\psi_1,\psi_3) + (\psi_1,\psi_4) + (\psi_2,\psi_3) + (\psi_2,\psi_4)$ $\lvert(\psi_1,\psi_2)\rvert \leqq \sqrt{(\psi_1,\psi_1)}\sqrt{(\psi_2,\psi_2)}$
Orthogonal vectors	$\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$	$(\psi_1,\psi_2) = 0$
Orthonormal basis set[3]	$\{\mathbf{e}_i\}$, with $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$ and also, for any $\mathbf{v}$ in $\mathcal{E}_3$, $\mathbf{v} = \sum_{i=1}^{3} (\mathbf{e}_i \cdot \mathbf{v})\mathbf{e}_i$	$\{\epsilon_i(x)\}$, with $(\epsilon_i,\epsilon_j) = \delta_{ij}$ and also, for any $\psi(x)$ in $\mathcal{H}$ $\psi(x) = \sum_{i=1}^{\infty} (\epsilon_i,\psi)\epsilon_i(x)$

[1] The inner product of two vectors is a *scalar*.
[2] The norm of a vector is a *nonnegative real number*.
[3] The scalars $\{\mathbf{e}_i \cdot \mathbf{v}\}$ and $\{(\epsilon_i, \psi)\}$ are called the *expansion coefficients* or *components* of the vectors $\mathbf{v}$ and $\psi(x)$ relative to the respective bases.

2-4 HILBERT SPACE OPERATORS

We recall from elementary calculus that a "function" f is, by definition, a "rule" which associates with each number x another number $y = f(x)$. This concept of a function can be extended to apply to vectors as well as numbers; however, it is then customary to use the term "operator" instead of "function." Thus, $\hat{O}$ is said to be an *operator* in the Hilbert space if and only if $\hat{O}$ specifies some rule of correspondence which associates with each vector $\psi(x)$ in $\mathcal{H}$ another vector $\phi(x)$. We write this as

$$\phi(x) = \hat{O}\psi(x) \tag{2-41}$$

and we speak of $\hat{O}$ as "operating on the vector $\psi(x)$, transforming it into the vector $\phi(x)$."

In the preceding section, we discussed the way in which a *vector* in $\mathcal{H}$ can be (i) multiplied by a scalar, (ii) added to another vector, and (iii) multiplied by another vector. Analogous operations can be defined for *operators* as well. Thus (i) the operator $c\hat{O}$ transforms a given vector $\psi(x)$ into the vector $c(\hat{O}\psi(x))$, (ii) the operator $\hat{O}_1 + \hat{O}_2$ transforms a given vector $\psi(x)$ into the vector $\hat{O}_1\psi(x) + \hat{O}_2\psi(x)$, and (iii) the operator $\hat{O}_1\hat{O}_2$ transforms a given vector $\psi(x)$ into the vector $\hat{O}_1(\hat{O}_2\psi(x))$. Thus, the product of c times $\hat{O}$, the sum of $\hat{O}_1$ and $\hat{O}_2$, and the product of $\hat{O}_1$ times $\hat{O}_2$, are *by definition* such that the following equations are valid for all vectors $\psi(x)$ in $\mathcal{H}$:

$$(c\hat{O})\psi(x) = c(\hat{O}\psi(x)) \tag{2-42a}$$

$$(\hat{O}_1 + \hat{O}_2)\psi(x) = \hat{O}_1\psi(x) + \hat{O}_2\psi(x) \tag{2-42b}$$

$$(\hat{O}_1\hat{O}_2)\psi(x) = \hat{O}_1(\hat{O}_2\psi(x)) \tag{2-42c}$$

In connection with Eq. (2-42c), it must be emphasized that it is *not* necessarily true that $\hat{O}_1\hat{O}_2 = \hat{O}_2\hat{O}_1$; in other words, it is not generally true that, for every $\mathcal{H}$-vector $\psi(x)$, $\hat{O}_1$ acting on $\hat{O}_2\psi(x)$ produces the same vector as does $\hat{O}_2$ acting on $\hat{O}_1\psi(x)$. *If*, however, the equality *does* hold for *all* vectors $\psi(x)$, then we say that $\hat{O}_1$ and $\hat{O}_2$ *commute*:

$\hat{O}_1$ and $\hat{O}_2$ *commute* if and only if $\hat{O}_1\hat{O}_2\psi(x) = \hat{O}_2\hat{O}_1\psi(x)$ for all vectors $\psi(x)$ in $\mathcal{H}$ (2-43)

Exercise 14. Let $\hat{O}_1 = x$ [i.e., $\hat{O}_1\psi(x) = x\psi(x)$], and let $\hat{O}_2 = d/dx$ [i.e., $\hat{O}_2\psi(x) = d\psi/dx$]. Show that $\hat{O}_1$ and $\hat{O}_2$ do *not* commute.

In quantum mechanics virtually all operators of interest possess a property called "linearity." By definition, an operator $\hat{O}$ is said to be a *linear* operator if and only if, for any $\mathcal{H}$-vectors $\psi_1(x)$ and

$\psi_2(x)$ and any complex numbers c_1 and c_2,

$$\hat{O}(c_1\psi_1(x) + c_2\psi_2(x)) = c_1\hat{O}\psi_1(x) + c_2\hat{O}\psi_2(x) \qquad (2\text{-}44)$$

Exercise 15. Show that the operator "d/dx" *is* a linear operator, but that the operator "log" is *not* a linear operator.

It is easy to show from Eqs. (2-42) that if $\hat{O}_1$ and $\hat{O}_2$ are linear operators, then the operators $(c_1\hat{O}_1 + c_2\hat{O}_2)$ and $\hat{O}_1\hat{O}_2$ are also linear.

In addition to linearity, another property which many operators in quantum mechanics possess is the property of "hermiticity." An operator $\hat{O}$ is said to be an *Hermitian* operator if and only if, for *any* two $\mathcal{H}$-vectors $\psi_1(x)$ and $\psi_2(x)$,

$$(\psi_1, \hat{O}\psi_2) = (\hat{O}\psi_1, \psi_2) \qquad (2\text{-}45)$$

As an example, let us see if the simple operator $\hat{O} = c$ is an Hermitian operator. If $\psi_1(x)$ and $\psi_2(x)$ are any two $\mathcal{H}$-vectors, then, using Eq. (2-34b), we have

$$(\psi_1, c\psi_2) = c(\psi_1, \psi_2) = (c^*\psi_1, \psi_2)$$

Thus, according to Eq. (2-45), the operator $\hat{O} = c$ is Hermitian if and only if $c = c^*$—i.e., if and only if c is real.

Exercise 16. If $\hat{O}_1$ and $\hat{O}_2$ are Hermitian operators, prove that

(a) $c_1\hat{O}_1 + c_2\hat{O}_2$ is Hermitian if c_1 and c_2 are real.

(b) $\hat{O}_1\hat{O}_2$ is Hermitian if $\hat{O}_1$ and $\hat{O}_2$ commute.

We turn now to discuss one final aspect of operators which will prove to be very essential to the mathematical formulation of quantum mechanics. We make the following definition: If the effect of a given operator $\hat{O}$ on some *particular* $\mathcal{H}$-vector $\psi(x)$ is to simply multiply that vector by an $\mathcal{H}$-scalar c,

$$\hat{O}\psi(x) = c\psi(x) \qquad (2\text{-}46)$$

then we say that the vector $\psi(x)$ is an *eigenvector* (or eigenfunction) of $\hat{O}$, and c is the corresponding *eigenvalue.*†

Exercise 17.

(a) Show that the function e^{ax} (where a is real) is an eigenfunction of the operator "d/dx." What is the corresponding eigenvalue?

(b) Show that the function x^n (where $n \geq 1$) is an eigenfunc-

† The prefix "eigen" is a German word. When we call c an *eigenvalue* of $\hat{O}$, we mean literally that c is a value which is "characteristic of" or "peculiar to" the operator $\hat{O}$.

tion of the operator "$x \cdot d/dx$." What is the corresponding eigenvalue?

(c) Of what operator (excluding $\hat{O} = c$) is the function $\cos ax$ an eigenfunction?

We can now establish two important results concerning the eigenvectors and eigenvalues of *Hermitian* operators:

(i) The eigenvalues of an Hermitian operator are pure real. To see this, suppose $\hat{O}$ is an Hermitian operator with eigenvector $\psi(x)$ and eigenvalue c. By Eqs. (2-46) and (2-34b), we can write

$$(\psi, \hat{O}\psi) = (\psi, c\psi) = c(\psi, \psi)$$

and

$$(\hat{O}\psi, \psi) = (c\psi, \psi) = c^*(\psi, \psi)$$

But, since $\hat{O}$ is Hermitian, then these two quantities must be equal:

$$c(\psi, \psi) = c^*(\psi, \psi)$$

Excluding the trivial case in which $\psi(x)$ is the null vector, we have $(\psi, \psi) > 0$, so we may conclude that $c = c^*$—i.e., c is pure real.

(ii) The eigenvectors corresponding to two unequal eigenvalues of an Hermitian operator are orthogonal to each other. The proof of this statement is the subject of the following exercise.

Exercise 18. Let $\hat{O}$ be an Hermitian operator with eigenfunctions $\psi_1(x)$ and $\psi_2(x)$, and let the corresponding eigenvalues c_1 and c_2 be unequal. Prove that $\psi_1(x)$ and $\psi_2(x)$ are orthogonal. [*Hint:* Consider the two quantities $(\psi_1, \hat{O}\psi_2)$ and $(\hat{O}\psi_1, \psi_2)$, and use the fact just established that c_1 and c_2 must be pure real.]

We shall now prove a theorem that is almost, but not quite, the converse of the preceding two theorems, (i) and (ii). Suppose $\hat{A}$ is a linear operator which possesses a *complete, orthonormal set of eigenvectors* $\{\alpha_n(x)\}$ and a corresponding *set of real eigenvalues* $\{a_n\}$:

$$\left.\begin{aligned} \hat{A}\alpha_n(x) &= a_n\alpha_n(x), \quad a_n \text{ real} \\ (\alpha_m, \alpha_n) &= \delta_{mn}, \quad \text{all } m,n \end{aligned}\right\} \tag{2-47}$$

We shall prove that these conditions imply that the operator $\hat{A}$ is *Hermitian*. We shall do this by showing that, for any two $\mathcal{H}$-vectors $\psi(x)$ and $\phi(x)$, it is true that $(\psi, \hat{A}\phi) = (\hat{A}\psi, \phi)$.

Define $c_n = (\alpha_n, \psi)$ and $e_n = (\alpha_n, \phi)$, and expand $\psi(x)$ and $\phi(x)$

in the orthonormal basis $\{\alpha_n(x)\}$ according to Eqs. (2-39):

$$\psi(x) = \sum_n c_n \alpha_n(x) \qquad \phi(x) = \sum_n e_n \alpha_n(x)$$

Using the linearity of $\hat{A}$, the properties of the inner product in Eqs. (2-34), and the orthonormality of the set $\{\alpha_n(x)\}$, we have

$$\begin{aligned}
(\hat{A}\psi, \phi) &= (\hat{A} \sum_n c_n \alpha_n, \sum_m e_m \alpha_m) \\
&= \left(\sum_n c_n A\alpha_n, \sum_m e_m \alpha_m \right) \\
&= \left(\sum_n c_n a_n \alpha_n, \sum_m e_m \alpha_m \right) \\
&= \sum_{m,n} c_n^* a_n^* e_m (\alpha_n, \alpha_m) \\
&= \sum_{m,n} c_n^* a_n^* e_m \delta_{nm}
\end{aligned}$$

so

$$(\hat{A}\psi, \phi) = \sum_n c_n^* a_n^* e_n$$

In an exactly analogous way, we find that

$$(\psi, \hat{A}\phi) = \sum_n c_n^* e_n a_n$$

Exercise 19. Carry out the steps leading to this last equation.

Since the eigenvalues $\{a_n\}$ were given to be real, $a_n^* = a_n$, it follows at once that $(\hat{A}\psi, \phi) = (\psi, \hat{A}\phi)$; therefore, $\hat{A}$ is Hermitian.

The foregoing theorem and its proof should be studied in detail. Not only is its content important, but the steps in its proof illustrate well the sort of mathematical manipulations which will be required in our development of quantum mechanics.

Exercise 20. Suppose a linear operator $\hat{A}$ has a complete, orthonormal set of eigenvectors $\{\alpha_n(x)\}$ and a corresponding set of

eigenvalues $\{a_n\}$. Show that a knowledge of the elements of the two sets $\{\alpha_n(x)\}$ and $\{a_n\}$ is sufficient to completely specify the operator. [*Hint:* Show that, for any *given* $\mathcal{H}$-vector $\psi(x)$, the vector $\hat{A}\psi(x)$ is completely defined through the quantities $\{\alpha_n(x)\}$ and $\{a_n\}$.]

CHAPTER

3

A Brief Review of Classical Mechanics

An important prerequisite for a meaningful understanding of quantum mechanics is a clear appreciation of the fundamental principles of classical mechanics. It is assumed that the reader is already familiar with the more elementary ideas and attitudes of classical mechanics. In this chapter we shall simply try to *organize*, and occasionally to *expand*, these ideas and attitudes, in a way that will best enable us to see later the basic similarities and differences between classical and quantum mechanics. At the same time, the level and approach of our development of the classical theory in this chapter should give the reader a rough indication of the level and approach of our development of the quantum theory in the next chapter.

3-1 A MECHANICAL SYSTEM

Our objects of study will be "mechanical systems" which, for the sake of simplicity, have only one degree of freedom. For concreteness, we shall take as our system a single particle of constant mass m which is constrained to move along the x-axis in a "conservative" force field, $F(x)$. In order to avoid the complications of the theory of relativity, we restrict our discussion to those cases in which the velocity of the particle,

$$v \equiv \frac{dx}{dt} \tag{3-1}$$

is always much less than the velocity of light.

The *force function* $F(x)$ gives the force exerted on the particle by its environment at each point x; thus the force function may be said to describe the mechanical interaction of the particle with its

environment. This interaction may also be described by the *potential function* $V(x)$, which *by definition* is that function whose negative derivative with respect to x is the force function:

$$F(x) \equiv -\frac{d}{dx} V(x) \tag{3-2}$$

Although it is possible to think of forces which cannot be derived from any function $V(x)$ according to Eq. (3-2) (e.g., a *frictional* force, which depends not on x, but rather on the direction of motion), our earlier stipulation that the force field be "conservative" means precisely that $V(x)$, as defined by Eq. (3-2), *does* exist. Consequently, for the systems of interest to us it makes no difference whatsoever whether we describe the interaction between the particle and its environment by specifying $F(x)$ or $V(x)$, since if one of these functions is known the other one may be found through Eq. (3-2).

The physical significance of the potential function may be understood as follows: Consider the graph of $V(x)$ in a small neighborhood of some point x_0. If the graph is sloping *upward* in this neighborhood, then $dV/dx|_{x_0} > 0$, and so Eq. (3-2) implies that the force $F(x_0)$ is in the *negative* x-direction; on the other hand, if the graph is sloping *downward*, then $dV/dx|_{x_0} < 0$, and Eq. (3-2) implies that the force $F(x_0)$ is in the *positive* x-direction. Thus, the force $F(x)$ always *tries* to move the particle in that direction which would result in a *decrease* in $V(x)$; moreover, Eq. (3-2) says that the strength or *magnitude* of this force at a given point is numerically equal to the *rate of decrease* of $V(x)$ at that point. We may therefore think of the graph of $V(x)$ as being a sort of "hilly terrain" upon which the particle rolls under the influence of some pseudogravitational force.

Exercise 21.

(a) Show that two potential functions which differ by only a constant (i.e., $V_2(x) = V_1(x) + C$) give rise to identical force functions, and hence are "physically equivalent" potentials.

(b) For the force field $F(x) = -kx$, what is the potential field? For the potential field $V(x) = k/x$, what is the force field?

(c) If the point x_0 is a *local minimum* of $V(x)$, show that x_0 is a point of *stable equilibrium*; i.e., show that the particle feels *no* force at the point x_0, while at any point slightly above or below x_0 the particle feels a force acting *toward* x_0. In a similar way, show that if x_0 is a *local maximum* of $V(x)$, then x_0 is a point of *unstable equilibrium*.

The basic program of mechanics, both classical and quantum, is essentially twofold: First, we have to decide how we shall specify the *instantaneous state* of a given mechanical system, and then we must

discover how this state changes or *evolves with time*. In this chapter we shall review how classical mechanics accomplishes these ends, and in the next chapter we shall consider the approach taken by quantum mechanics.

3-2 THE CLASSICAL STATE

In classical mechanics the instantaneous state of a mechanical system is described in terms of the values of certain "observable variables" of the system. For the simple system of a particle of mass m constrained to move along the x-axis, the observable variables used to define the state are normally the *position* x and the *momentum* $p \equiv mv$ of the particle. In other words, the state of the system at time t is specified by the pair of values $[x(t),p(t)]$.

This classical definition of the state of a mechanical system tacitly assumes that:

(i) the position and momentum variables both have precise, well-defined values at each instant of time; and

(ii) it is always possible, at least in principle, to measure these values without significantly disturbing the system.

These assumptions might seem so natural and innocuous as to hardly merit mentioning. However, we shall find in the next chapter that quantum mechanics, in its most widely accepted formulation, actually *denies* the general validity of *both* these assumptions! This comes about as a consequence of the radically different viewpoint taken by quantum mechanics with regard to the concepts of "state" and "observable variables," and to the role played by the measurement process. An elucidation of these important points will occupy much of our attention in the following chapter.

3-3 TIME EVOLUTION OF THE CLASSICAL STATE

3-3a The Newtonian Formulation

Having defined the instantaneous state of our classical, one-particle system, we must now address ourselves to the problem of discovering how the state changes with time. The assumption of course is that the state variables x and p stand in a definite functional relationship to the time variable t, and our object is to de-

termine the precise *forms* of the functions $x(t)$ and $p(t)$. One way to achieve these ends is to *postulate* Newton's second law,

$$\text{Force} = \text{mass} \times \text{acceleration}$$

which we shall write here in the form

$$\frac{d^2x}{dt^2} = \frac{F(x)}{m} \tag{3-3a}$$

In words, this law says that the desired function $x(t)$ is that function whose second time derivative is equal to the force function divided by the particle mass. Evidently, then, to find $x(t)$ we must integrate Eq. (3-3a) twice with respect to t; once $x(t)$ is found, we then obtain the function $p(t)$ by differentiating $x(t)$ and multiplying by m:

$$p = m\frac{dx}{dt} \tag{3-3b}$$

To be sure, the twofold integration of Newton's second law may be very difficult, if not impossible, to perform analytically for certain force functions $F(x)$. However, this is more a problem of applied mathematics rather than physics (although it sometimes happens that a "feeling" for a given physical problem will suggest a fruitful way of carrying out the requisite integrations). For our purposes, though, we merely content ourselves with the thought that Eq. (3-3a) *in principle* determines the function $x(t)$ for any given force field $F(x)$, regardless of how difficult it may be to explicitly *solve* the differential equation.

In twice integrating Eq. (3-3a), we will generate two constants of integration. The values of these two constants may be uniquely fixed by specifying the values of x and dx/dt at some "initial time" $t = 0$. Equivalently, since dx/dt and p are related by Eq. (3-3b), we may specify the initial value of p instead of dx/dt.

In conclusion we see that, with either $F(x)$ or $V(x)$ given, and with the "initial state" $[x(0), p(0)]$ specified, then Eqs. (3-3a) and (3-3b) enable us to determine unambiguously the state of the system $[x(t),p(t)]$ at any time $t > 0$.†

†If the reader wondered why, in the previous section, we chose to specify the state of our system by the *pair* of variables $[x(t),p(t)]$, rather than by $x(t)$ alone, the reason should now be apparent: since the time evolution equation for $x(t)$, Eq. (3-3a), contains a *second*-order time derivative, then a specification of $x(0)$ alone would *not* suffice to unambiguously determine $x(t)$ for all $t > 0$. Generally speaking, the observable variables chosen to specify the "classical state" of a system must be such that their initial values collectively determine their subsequent values.

As a familiar example, for the simple force field $F(x) = k$, or $V(x) = -kx$, Newton's second law reads

$$\frac{d^2x}{dt^2} = \frac{k}{m}$$

The two t-integrations yield successively

$$\frac{dx}{dt} = \frac{k}{m}t + C_1, \qquad x = \frac{1}{2}\frac{k}{m}t^2 + C_1 t + C_2$$

Hence,

$$x(t) = \frac{1}{2}\frac{k}{m}t^2 + C_1 t + C_2, \qquad p(t) = m\frac{dx}{dt} = kt + mC_1$$

The requirements that $x(0) = x_0$ and $p(0) = p_0$ imply that $C_2 = x_0$ and $mC_1 = p_0$. Consequently, if $[x_0, p_0]$ is the state at time 0, then the state at time t is evidently

$$\left[\frac{1}{2}\frac{k}{m}t^2 + \frac{p_0}{m}t + x_0, \qquad kt + p_0\right]$$

3-3b Energy

One very important consequence of Newton's second law is the introduction of the concept of energy. This concept arises naturally through a consideration of the quantity

$$W_{12} \equiv \int_{x_1}^{x_2} F(x)dx \tag{3-4}$$

which is called "the work done on the particle by the force $F(x)$ during the motion from x_1 to x_2." Of course this definition, like any definition, tells us nothing new; however, let us use the two expressions for $F(x)$ given in Eqs. (3-2) and (3-3a) to calculate two different expressions for W_{12}, and then see what we can learn by equating the results. First, from the definition of the potential function in Eq. (3-2), we have

$$W_{12} = \int_{x_1}^{x_2} \left(-\frac{dV}{dx}\right)dx = -\int_{x_1}^{x_2} dV = -V(x)\Big|_{x_1}^{x_2}$$

so

$$W_{12} = -[V(x_2) - V(x_1)] \tag{3-5a}$$

Second, from Newton's second law in Eq. (3-3a), making use of Eq. (3-1) and the chain rule for derivatives, we have

$$W_{12} = \int_{x_1}^{x_2} \left(m \frac{d^2x}{dt^2}\right) dx = \int_{x_1}^{x_2} \left(m \frac{dv}{dt}\right) dx = \int_{x_1}^{x_2} \left(m \frac{dv}{dx} \frac{dx}{dt}\right) dx$$

$$= \int_{x_1}^{x_2} \left(mv \frac{dv}{dx}\right) dx = \int_{v_1}^{v_2} mv dv = \frac{1}{2} mv^2 \Big|_{v_1}^{v_2}$$

so

$$W_{12} = \frac{1}{2} mv_2^2 - \frac{1}{2} mv_1^2 \tag{3-5b}$$

where v_1 and v_2 are the velocities of the particle at x_1 and x_2 respectively. Equating the two expressions for W_{12} in Eqs. (3-5a) and (3-5b), we find that

$$\frac{1}{2} mv_1^2 + V(x_1) = \frac{1}{2} mv_2^2 + V(x_2) \tag{3-5c}$$

But since the points x_1 and x_2 were completely arbitrary, then we may conclude that the quantity

$$E \equiv \frac{1}{2} mv^2 + V(x) \tag{3-6a}$$

maintains a constant value throughout the motion of the particle. E is called the *energy* of the system. It is seen to be composed of two parts, one part associated with the *motion*, $mv^2/2$, and the other part associated with the *position*, $V(x)$; evidently, the particle moves in such a way that any decrease in one of these terms is always exactly compensated by an increase in the other term.

To demonstrate the usefulness of the energy concept in studying the dynamics of a mechanical system, we show in Fig. 2 a plot of $V(x)$ versus x for a particle in some hypothetical force field. If the total energy is E, then the particle is constrained to move only in those regions in which $V(x) < E$, since the quantity $mv^2/2$ in Eq. (3-6a) cannot go negative. In such a region, the distance from the curve $V(x)$ up to the line $V = E$ is evidently just the difference between the *total* energy E and the *potential* energy $V(x)$—i.e., this distance is just the *kinetic* energy $mv^2/2$. In Fig. 2 we have illustrated how the energy of the particle at a given point is divided between the kinetic and potential terms. The points $x = a$ and $x = b$ satisfy $V(a) = V(b) = E$, and define the *boundaries* of the motion; at these points the velocity of the particle must obviously vanish, and

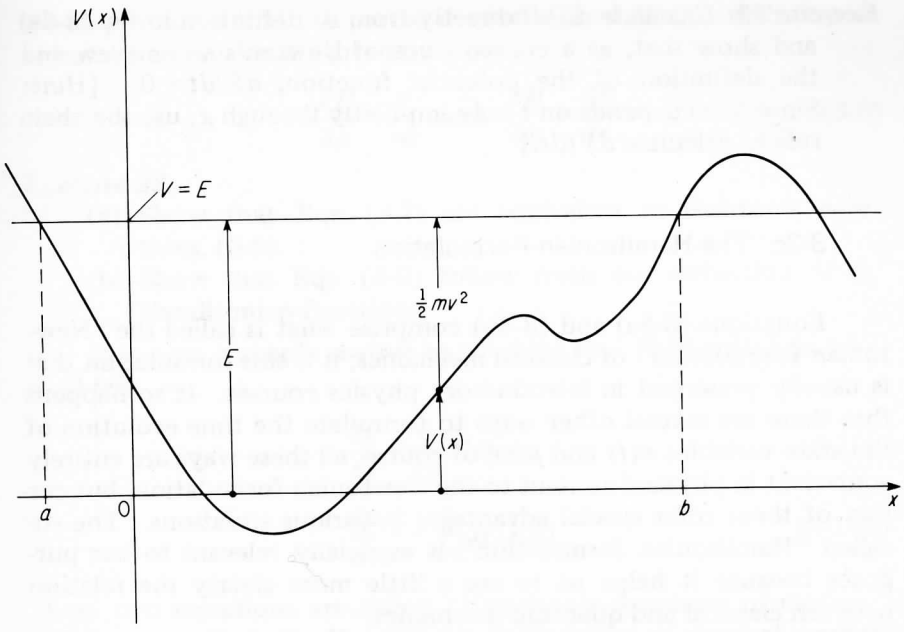

Fig. 2. A plot of the potential function $V(x)$ versus x for a particle in some hypothetical force field. The distance from the curve $V(x)$ up to the horizontal line $V = E$ represents the kinetic energy, $\frac{1}{2}mv^2$. Since this quantity cannot become negative, the motion of the particle is confined to the interval $a < x < b$, where $V(x) < E$. The distance from the base line up to the curve represents the potential energy $V(x)$, which can be either positive or negative.

the particle will evidently reverse its direction of travel. Thus the energy plot in Fig. 2 indeed presents a very clear and concise picture of the motion of the particle: the distance $(E - V(x))$ provides a measure of the *speed* of the particle at the point x, while the slope of the potential function curve at the point x provides a measure of the direction and magnitude of the *force* on the particle, or the *acceleration* of the particle, at this point.

Since $p = mv$, then we can also write E as

$$E = \frac{p^2}{2m} + V(x) \tag{3-6b}$$

The fact that *this particular function* of the state variables x and p remains *constant* as these two variables evolve with time is clearly a non-trivial consequence of Newton's second law, and is why the concept of energy occupies such an important place in the structure of classical mechanics.

Exercise 22. Calculate dE/dt directly from its definition in Eq. (3-6a) and show that, as a consequence of Newton's second law and the definition of the potential function, $dE/dt = 0$. [*Hint*: Since $V(x)$ depends on t only implicitly through x, use the chain rule to calculate dV/dt.]

3-3c The Hamiltonian Formulation

Equations (3-3a) and (3-3b) comprise what is called the "Newtonian formulation" of classical mechanics; it is this formulation that is usually presented in introductory physics courses. It so happens that there are several other ways to formulate the time evolution of the state variables $x(t)$ and $p(t)$; of course, all these ways are entirely equivalent in physical content to the Newtonian formulation, but certain of them offer special advantages in various situations. The so-called "Hamiltonian formulation" is especially relevant to our purposes because it helps us to see a little more clearly the relation between classical and quantum mechanics.

From a strictly classical point of view, the Hamiltonian formulation is nice because it places the state variables x and p on a formally equivalent footing. We note that, by contrast, the Newtonian formulation treats x as the primary variable [see Eq. (3-3a)], while p seems to be merely an auxiliary variable which is derived from x [see Eq. (3-3b)]. For the particular case of a particle on the x-axis in a potential field $V(x)$, the transition from the Newtonian to the Hamiltonian formulation can be easily made: Using Eqs. (3-1) and (3-2), we can write Eqs. (3-3b) and (3-3a) respectively as

$$\frac{dx}{dt} = \frac{p}{m} \tag{3-7a}$$

and

$$\frac{dp}{dt} = -\frac{dV}{dx} \tag{3-7b}$$

We now *define* the *Hamiltonian function* $H(x,p)$ to be the total energy of the system expressed as a function of the state variables x and p. Thus, using Eq. (3-6b), we define

$$H(x,p) \equiv \frac{p^2}{2m} + V(x) \tag{3-8}$$

We now observe that the partial derivatives† of $H(x,p)$ with respect to its two arguments are given by

$$\frac{\partial H}{\partial p} = \frac{p}{m} \quad \text{and} \quad \frac{\partial H}{\partial x} = \frac{dV}{dx} \tag{3-9}$$

Exercise 23.

(a) Show that Eqs. (3-7) are equivalent to Newton's equations, (3-3).

(b) Show that Eqs. (3-9) follow from our definition of the Hamiltonian function.

It is now a simple matter to combine Eqs. (3-9) with Eqs. (3-7) to obtain

$$\frac{dx}{dt} = \frac{\partial}{\partial p} H(x,p) \tag{3-10a}$$

$$\frac{dp}{dt} = -\frac{\partial}{\partial x} H(x,p) \tag{3-10b}$$

These two equations are called "Hamilton's equations of motion," and they constitute the Hamiltonian formulation of classical mechanics in the same way that Eqs. (3-3a) and (3-3b) constitute the Newtonian formulation. It should be clear from our derivation of Hamilton's equations that they contain no more nor no less information than Newton's equations. However, it is equally clear that the state variables x and p *formally* play more symmetric roles in Eqs. (3-10) than in Eqs. (3-3). From a strictly mathematical point of view, what we have done is to replace a single, second-order differential equation in the one variable x [Eq. (3-3a)] with two coupled, first-order differential equations in the two variables x and p [Eqs. (3-10)]. It should also be noted that, in the Hamiltonian formulation, the mechanical interaction of the particle with its environment is now formally described by the Hamiltonian function, rather than by the force function or the potential function. Of course, it is clear from the definitions in Eqs. (3-2) and (3-8) that a knowledge of any one of the functions $H(x,p)$, $V(x)$, or $F(x)$, implies a knowledge of the other two.

To find the functions $x(t)$ and $p(t)$ via the Hamiltonian formu-

†If f is a function of two variables u and v, then $\partial f/\partial u$, the "partial derivative of f with respect to u," is defined to be the u-derivative taken with v treated as a constant. Thus if $f(u,v) = u^2v^3$, then we have $\partial f/\partial u = 2uv^3$, and $\partial f/\partial v = 3v^2u^2$.

lation, we must evidently integrate Eqs. (3-10a) and (3-10b) once each with respect to t. Again, the actual integrations might be quite difficult to *perform*—especially since the equations are *coupled* (i.e., x and p both appear in both equations); however, we are not concerned here with these "practical details." Of more interest to us is the fact that the two integrations will yield two integration constants, and, as in the Newtonian formulation, these two constants can be fixed by specifying the values of x and p at some "initial time" $t = 0$. Thus, with $H(x,p)$ given, and with the initial state $[x(0),p(0)]$ specified, then Eqs. (3-10a) and (3-10b) enable us to determine unambiguously the state of the system $[x(t),p(t)]$ at any time $t > 0$.

Exercise 24. We have derived Hamilton's equations from Newton's equations. Prove now that Newton's equations can be derived from Hamilton's equations. That is, show that Eqs. (3-3) follow from Eqs. (3-10) when account is taken of the definitions of $H(x,p)$ and $V(x)$.

3-3d "Determinism" in Classical Mechanics

Classical mechanics has sometimes been said by philosophers to imply a "deterministic" universe. By this it is meant that, given the initial state of the universe (i.e., given the exact positions and momenta of all the particles in the universe at some time $t = 0$), and given also the functional forms of all the forces acting on and among these particles, then the subsequent "history" of the universe is in principle completely *determined* through the dynamical equations of Newton (or Hamilton). We shall see in the next chapter that the tenets of quantum mechanics will force us to dramatically revise (but not to completely discard) this simple deterministic picture of the physical universe.