A Shorter Course of Theoretical Physics

IN THREE VOLUMES

Vol. 1. Mechanics and Electrodynamics
Vol. 2. Quantum Mechanics
Vol. 3. Macroscopic Physics

A SHORTER COURSE OF THEORETICAL PHYSICS

VOLUME 2

QUANTUM MECHANICS

BY

L. D. LANDAU

AND

E. M. LIFSHITZ

Institute of Physical Problems, U.S.S.R. Academy of Sciences

TRANSLATED FROM THE RUSSIAN BY
J. B. SYKES AND J. S. BELL

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY

Pergamon Press Ltd., Headington Hill Hall, Oxford
Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523
Pergamon of Canada Ltd., 207 Queen's Quay West, Toronto 1
Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia

---

First edition 1974

Library of Congress Cataloging in Publication Data

Landau, Lev Davidovich, 1908–1968.
A shorter course of theoretical physics.
Translation of Kratkiĭ kurs teoreticheskoĭ fiziki.
CONTENTS: v. 1. Mechanics and electrodynamics.
—v. 2. Quantum mechanics.
1. Physics. 2. Mechanics. 3. Quantum theory.
I. Lifshits, Evgeniĭ Mikhaĭlovich, joint author.
II. Title.
QC21.2.L3513 530 74–167927
ISBN 0–08–016739–X (v. 1)
ISBN 0–08–017801–4 (v. 2)

Translated from *Kratkiĭ kurs teoreticheskoĭ fiziki, Kniga 2: Kvantovaya Mekhanika*
Izdatel'stvo "Nauka", Moscow, 1972

Printed in Hungary

PREFACE

THIS book continues with the plan originated by Lev Davidovich Landau and described in the Preface to Volume 1: to present the minimum of material in theoretical physics that should be familiar to every present-day physicist, working in no matter what branch of physics.

Part I, dealing with non-relativistic quantum theory, follows our *Quantum Mechanics* (Volume 3 of the *Course of Theoretical Physics*). This has been abridged by dropping completely some sections that are of interest only to specialists, as well as numerous details of technique that are intended for those whose profession lies in theoretical physics. This considerable abridgement has naturally meant rewriting a fairly large part of the book. I have nevertheless tried to keep unchanged the manner and style of the exposition, and in no place to allow a simplification by popularising; the only simplification is by the omission of detail. In Part I, the words "it can be shown" hardly occur: the results given are accompanied by their derivations.

This is, however, less true of Part II. The treatment here is based on the *Relativistic Quantum Theory* by Berestetskiĭ, Pitaevskiĭ and myself (Volume 4 of the *Course*), but only the fundamentals of quantum electrodynamics are presented. Here again I have sought to proceed in such a way as to show as clearly as possible the physical hypotheses and logical structure of the theory; but many applications of the theory are mentioned only by way of their results, on account of the frequent complexity of the calculations needed to solve specific problems in this field. In the choice of materials for Part II I have also been guided to some extent by the content of Landau's lectures on quantum electrodynamics at Moscow University in 1959–60; my thanks are due to A. S. Kompaneets, N. I. Bud'ko and P. S. Kondratenko for making available their notes of these lectures.

The final chapter on Feynman diagrams differs somewhat in style, both as regards its greater complexity and in being concerned with methods rather than physical results. I felt it necessary, however, to provide the reader with at least an idea of the origin and significance of the concepts of the diagram technique, which are an indispensable part of the equipment of theoretical physics at the present time. (I do not seek to describe the use of this technique for the solution of practical problems.) This chapter can be omitted, if the reader so wishes, without affecting the study of the remainder.

This book was published in the original Russian almost exactly ten years after the fateful day of 7 January 1962, when a road accident cut short Lev Davidovich Landau's work as a scientist and a teacher. Not one of the readers of the *Shorter Course* has had the joy of attending Landau's lectures. I should like to think that in these books it will be possible to convey to them something of his spirit as a teacher, his striving for clarity, his effort to make simple what was complex and so to reveal the laws of nature in their true simplicity and beauty.

E. M. Lifshitz

PUBLISHER'S NOTE

As IS the general rule in the volumes in the *Course of Theoretical Physics*, references to original papers give simply the author's name and the date.

on quantum electrodynamics at Moscow University in 1959–60; my thanks are due to A. S. Kompaneets, N. I. Bud'ko and P. S. Kondratenko for making available their notes of these lectures.

The final chapter on Feynman diagrams differs somewhat in style, both as regards its greater complexity and in being concerned with methods rather than physical results. I felt it necessary, however, to provide the reader with at least an idea of the origin and significance of the concepts of the diagram technique, which are an indispensable part of the equipment of theoretical physics at the present time. (I do not seek to describe the use of this technique for the solution of practical problems.) This chapter can be omitted, if the reader so wishes, without affecting the study of the remainder.

This book was published in the original Russian almost exactly ten years after the fateful day of 7 January 1962, when a road accident cut short Lev Davidovich Landau's work as a scientist and a teacher. Not one of the readers of the *Shorter Course* has had the joy of attending Landau's lectures. I should like to think that in these books it will be possible to convey to them something of his spirit as a teacher, his striving for clarity, his effort to make simple what was complex and so to reveal the laws of nature in their true simplicity and beauty.

E. M. Lifshitz

PUBLISHER'S NOTE

As is the general rule in the volumes in the *Course of Theoretical Physics*, references to original papers give simply the author's name and the date.

NOTATION

Ψ time-dependent wave function

ψ wave function without time factor

Operators are denoted by a circumflex $\hat{}$

Transposed operators are denoted by a tilde $\tilde{}$

Hermitian conjugate operators are denoted by a superscript $^{+}$

$f_{mn} = \langle m|f|n\rangle$ matrix elements of the quantity f

$\hat{H}$ Hamiltonian

E non-relativistic energy

$\omega_{nm} = (E_n - E_m)/\hbar$ transition frequency

ε relativistic particle energy, including rest energy

$\mathrm{d}q$ element in configuration space

$\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ element in ordinary space

Ω normalisation volume

Four-dimensional vector indices are denoted (in Part II) by Greek letters λ, μ, ν, . . ., which take the values 0, 1, 2, 3.

In Part II, relativistic units are used; they are defined in the first footnote to §76.

References to *Mechanics and Electrodynamics* are to Volume 1 of the *Shorter Course.*

CHAPTER 1

THE BASIC CONCEPTS OF QUANTUM MECHANICS

§1. The uncertainty principle

When we attempt to apply classical mechanics and electrodynamics to explain atomic phenomena, they lead to results which are in obvious conflict with experiment. This is very clearly seen from the contradiction obtained on applying ordinary electrodynamics to a model of an atom in which the electrons move round the nucleus in classical orbits. During such motion, as in any accelerated motion of charges, the electrons would have to emit electromagnetic waves continually. By this emission, the electrons would lose their energy, and this would eventually cause them to fall into the nucleus. Thus, according to classical electrodynamics, the atom would be unstable, which does not at all agree with reality.

This marked contradiction between theory and experiment indicates that the construction of a theory applicable to atomic phenomena —that is, phenomena occurring in particles of very small mass at very small distances—demands a fundamental modification of the basic physical concepts and laws.

As a starting-point for an investigation of these modifications, it is convenient to take the experimentally observed phenomenon known as *electron diffraction*.[†] It is found that, when a homogeneous beam

---

[†] The phenomenon of electron diffraction was in fact discovered after quantum mechanics was invented. In our discussion, however, we shall not adhere to the historical sequence of development of the theory, but shall endeavour to construct it in such a way that the connection between the basic principles of quantum mechanics and the experimentally observed phenomena is most clearly shown.

of electrons passes through a crystal, the emergent beam exhibits a pattern of alternate maxima and minima of intensity, wholly similar to the diffraction pattern observed in the diffraction of electromagnetic waves. Thus, under certain conditions, the behaviour of material particles—in this case, the electrons—displays features belonging to wave processes.

How markedly this phenomenon contradicts the usual ideas of motion is best seen from the following imaginary experiment, an idealisation of the experiment of electron diffraction by a crystal. Let us imagine a screen impermeable to electrons, in which two slits are cut. On observing the passage of a beam of electrons through one of the slits, the other being covered, we obtain, on a continuous screen placed behind the slit, some pattern of intensity distribution; in the same way, by uncovering the second slit and covering the first, we obtain another pattern. On observing the passage of the beam through both slits, we should expect, on the basis of ordinary classical ideas, a pattern which is a simple superposition of the other two: each electron, moving in its path, passes through one of the slits and has no effect on the electrons passing through the other slit. The phenomenon of electron diffraction shows, however, that in reality we obtain a diffraction pattern which, owing to interference, does not at all correspond to the sum of the patterns given by each slit separately. It is clear that this result can in no way be reconciled with the idea that electrons move in paths.

Thus the mechanics which governs atomic phenomena —*quantum mechanics* or *wave mechanics*—must be based on ideas of motion which are fundamentally different from those of classical mechanics. In quantum mechanics there is no such concept as the path of a particle. This forms the content of what is called the *uncertainty principle*, one of the fundamental principles of quantum mechanics, discovered by W. Heisenberg in 1927.[†]

[†] It is of interest to note that the complete mathematical formalism of quantum mechanics was constructed by W. Heisenberg and E. Schrödinger in 1925–6, before the discovery of the uncertainty principle, which revealed the physical content of this formalism.

In that it rejects the ordinary ideas of classical mechanics, the uncertainty principle might be said to be negative in content. Of course, this principle in itself does not suffice as a basis on which to construct a new mechanics of particles. Such a theory must naturally be founded on some positive assertions, which we shall discuss below (§2). However, in order to formulate these assertions, we must first ascertain the statement of the problems which confront quantum mechanics. To do so, we first examine the special nature of the interrelation between quantum mechanics and classical mechanics. A more general theory can usually be formulated in a logically complete manner, independently of a less general theory which forms a limiting case of it. Thus, relativistic mechanics can be constructed on the basis of its own fundamental principles, without any reference to Newtonian mechanics. It is in principle impossible, however, to formulate the basic concepts of quantum mechanics without using classical mechanics. The fact that an electron[†] has no definite path means that it has also, in itself, no other dynamical characteristics.[‡] Hence it is clear that, for a system composed only of quantum objects, it would be entirely impossible to construct any logically independent mechanics. The possibility of a quantitative description of the motion of an electron requires the presence also of physical objects which obey classical mechanics to a sufficient degree of accuracy. If an electron interacts with such a "classical object", the state of the latter is, generally speaking, altered. The nature and magnitude of this change depend on the state of the electron, and therefore may serve to characterise it quantitatively.

In this connection the "classical object" is usually called *apparatus*, and its interaction with the electron is spoken of as *measurement*. However, it must be emphasised that we are here not discussing a

---

[†] In this and the following sections we shall, for brevity, speak of "an electron", meaning in general any object of a quantum nature, i.e. a particle or system of particles to which classical mechanics cannot be applied.

[‡] We refer to quantities which characterise the motion of the electron, and not to those, such as the charge and the mass, which relate to it as a particle; these are parameters.

process of measurement in which the physicist-observer takes part. By *measurement*, in quantum mechanics, we understand any process of interaction between classical and quantum objects, occurring apart from and independently of any observer. The importance of the concept of measurement in quantum mechanics was elucidated by N. Bohr.

We have defined "apparatus" as a physical object which is governed, with sufficient accuracy, by classical mechanics. Such, for instance, is a body of large enough mass. However, it must not be supposed that apparatus is necessarily macroscopic. Under certain conditions, the part of apparatus may also be taken by an object which is microscopic, since the idea of "with sufficient accuracy" depends on the actual problem proposed. Thus, the motion of an electron in a Wilson chamber is observed by means of the cloudy track which it leaves, and the thickness of this is large compared with atomic dimensions; when the path is determined with such low accuracy, the electron is an entirely classical object.

Thus quantum mechanics occupies a very unusual place among physical theories: it contains classical mechanics as a limiting case, yet at the same time it requires this limiting case for its own formulation.

We may now formulate the problem of quantum mechanics. A typical problem consists in predicting the result of a subsequent measurement from the known results of previous measurements. Moreover, we shall see later that, in comparison with classical mechanics, quantum mechanics, generally speaking, restricts the range of values which can be taken by various physical quantities (for example, energy): that is, the values which can be obtained as a result of measuring the quantity concerned. The methods of quantum mechanics must enable us to determine these admissible values.

The measuring process has in quantum mechanics a very important property: it always affects the electron subjected to it, and it is in principle impossible to make its effect arbitrarily small, for a given accuracy of measurement. The more exact the measurement, the stronger the effect exerted by it, and only in measurements of very low accuracy can the effect on the measured object be small. This property of measurements is logically related to the fact that the dynamical characteristics of the electron appear only as a result of the measurement itself. It is clear that, if the effect of the measuring process on the object of it could be made arbitrarily small, this would mean that the measured quantity has in itself a definite value independent of the measurement.

Among the various kinds of measurement, the measurement of the coordinates of the electron plays a fundamental part. Within the limits of applicability of quantum mechanics, a measurement of the coordinates of an electron can always be performed[†] with any desired accuracy.

Let us suppose that, at definite time intervals Δt, successive measurements of the coordinates of an electron are made. The results will not in general lie on a smooth curve. On the contrary, the more accurately the measurements are made, the more discontinuous and disorderly will be the variation of their results, in accordance with the non-existence of a path of the electron. A fairly smooth path is obtained only if the coordinates of the electron are measured with a low degree of accuracy, as for instance from the condensation of vapour droplets in a Wilson chamber.

If now, leaving the accuracy of the measurements unchanged, we diminish the intervals Δt between measurements, then adjacent measurements, of course, give neighbouring values of the coordinates. However, the results of a series of successive measurements, though they lie in a small region of space, will be distributed in this region in a wholly irregular manner, lying on no smooth curve.

This circumstance shows that, in quantum mechanics, there is no such concept as the velocity of a particle in the classical sense of the word, i.e. the limit to which the difference of the coordinates at two instants, divided by the interval Δt between these instants, tends as Δt tends to zero. However, we shall see later that in quantum mechanics,

---

[†] Once again we emphasise that, in speaking of "performing a measurement", we refer to the interaction of an electron with a classical "apparatus", which in no way presupposes the presence of an external observer.

nevertheless, a reasonable definition of the velocity of a particle at a given instant can be constructed, and this velocity passes into the classical velocity as we pass to classical mechanics. But whereas in classical mechanics a particle has defined coordinates and velocity at any given instant, in quantum mechanics the situation is entirely different. If, as a result of measurement, the electron is found to have definite coordinates, then it has no definite velocity whatever. Conversely, if the electron has a definite velocity, it cannot have a definite position in space. For the simultaneous existence of the coordinates and velocity would mean the existence of a definite path, which the electron has not. Thus, in quantum mechanics, the coordinates and velocity of an electron are quantities which cannot be simultaneously measured exactly, i.e. they cannot simultaneously have definite values. We may say that the coordinates and velocity of the electron are quantities which do not exist simultaneously. In what follows we shall derive the quantitative relation which determines the possibility of an inexact measurement of the coordinates and velocity at the same instant.

A complete description of the state of a physical system in classical mechanics is effected by stating all its coordinates and velocities at a given instant; with these initial data, the equations of motion completely determine the behaviour of the system at all subsequent instants. In quantum mechanics such a description is in principle impossible, since the coordinates and the corresponding velocities cannot exist simultaneously. Thus a description of the state of a quantum system is effected by means of a smaller number of quantities than in classical mechanics, i.e. it is less detailed than a classical description.

A very important consequence follows from this regarding the nature of the predictions made in quantum mechanics. Whereas a classical description suffices to predict the future motion of a mechanical system with complete accuracy, the less detailed description given in quantum mechanics evidently cannot be enough to do this. This means that, even if an electron is in a state described in the most complete manner possible in quantum mechanics, its behaviour at subsequent instants is still in principle uncertain. Hence quantum mechanics cannot make completely definite predictions concerning the future behaviour of the electron. For a given initial state of the electron, a subsequent measurement can give various results. The problem in quantum mechanics consists in determining the probability of obtaining various results on performing this measurement. It is understood, of course, that in some cases the probability of a given result of measurement may be equal to unity, i.e. certainty, so that the result of that measurement is unique.

We shall often find in what follows that by no means every set of physical quantities in quantum mechanics can be measured simultaneously, i.e. can all have definite values at the same time. We have already mentioned one example, namely the velocity and coordinates of an electron. An important part is played in quantum mechanics by sets of physical quantities having the following property: these quantities can be measured simultaneously, but if they simultaneously have definite values, no other physical quantity (not being a function of these) can have a definite value in that state. We shall speak of such sets of physical quantities as *complete sets*.

Any description of the state of an electron arises as a result of some measurement. We shall now formulate the meaning of a *complete description* of a state in quantum mechanics. Completely described states occur as a result of the simultaneous measurement of a complete set of physical quantities. From the results of such a measurement we can, in particular, determine the probability of various results of any subsequent measurement, regardless of the history of the electron prior to the first measurement.

From now on (except in §§7 and 42) we shall understand the states of a quantum system to be completely described states.

§2. The principle of superposition

The fundamental difference between the physical concepts of motion in quantum and classical mechanics naturally implies an equally fundamental change in the mathematical formalism of the theory. First of all, therefore, we must consider the method of describing the state of a quantum system.

We shall denote by q the set of coordinates of a quantum system, and by $\mathrm{d}q$ the product of the differentials of these coordinates. This $\mathrm{d}q$ is called an element of volume in the *configuration space* of the system; for one particle, $\mathrm{d}q$ coincides with an element of volume $\mathrm{d}V$ in ordinary space.

In classical mechanics, the state of a system is described by specifying, at a particular instant, all its coordinates q and velocities $\dot{q}$. In quantum mechanics, as we have seen, such a description is certainly impossible. A complete description of the state of the system here signifies much less, namely the possibility of predicting the probabilities of particular results when the coordinates (or other quantities) of the system are measured.

The basis of the mathematical formalism of quantum mechanics lies in the proposition that a state of a system can be described by a definite (in general complex) function $\Psi(q)$ of the coordinates. The square of the modulus of this function determines the probability distribution of the values of the coordinates: $|\Psi|^2\,\mathrm{d}q$ is the probability that a measurement performed on the system will find the values of the coordinates to be in the element $\mathrm{d}q$ of configuration space. The function Ψ is called the *wave function* of the system.[†]

A knowledge of the wave function allows us, in principle, to calculate the probability of the various results of any other measurement (not necessarily of the coordinates) also. All these probabilities are determined by expressions bilinear in Ψ and Ψ^*. The most general form of such an expression is

$$\iint \Psi(q)\,\Psi^*(q')\,\phi(q, q')\,\mathrm{d}q\,\mathrm{d}q', \tag{2.1}$$

where the function $\phi(q, q')$ depends on the nature and the result of the measurement, and the integration is extended over all configuration space. The probability $\Psi^*\Psi$ of various values of the coordinates is itself an expression of this type.

The state of the system, and with it the wave function, in general varies with time. In this sense the wave function can be regarded as a

---

[†] It was first introduced into quantum mechanics by Schrödinger in 1926.

function of time also. If the wave function is known at some initial instant, then, from the very meaning of the concept of complete description of a state, it is in principle determined at every succeeding instant. The actual dependence of the wave function on time is determined by equations which will be derived later.

The sum of the probabilities of all possible values of the coordinates of the system must, by definition, be equal to unity. It is therefore necessary that the result of integrating $|\Psi|^2$ over all configuration space should be equal to unity:

$$\int |\Psi|^2 \, dq = 1. \tag{2.2}$$

This equation is what is called the *normalisation condition* for wave functions. If the integral of $|\Psi|^2$ converges, then by choosing an appropriate constant coefficient the function Ψ can be, as we say, *normalised*. We shall see later, however, that the integral of $|\Psi|^2$ may diverge, and then Ψ cannot be normalised by the condition (2.2). In such cases $|\Psi|^2$ does not, of course, determine the absolute values of the probability of the coordinates, but the ratio of the values of $|\Psi|^2$ at two different points of configuration space determines the relative probability of the corresponding values of the coordinates.

Since all quantities calculated by means of the wave function, and having a direct physical meaning, are of the form (2.1), in which Ψ appears multiplied by Ψ^*, it is clear that the normalised wave function is determined only to within a constant *phase factor* of the form $e^{i\alpha}$ (where α is any real number). This indeterminacy is in principle irremovable; it is, however, unimportant, since it has no effect upon any physical results.

The positive content of quantum mechanics is founded on a series of propositions concerning the properties of the wave function. These are as follows.

Suppose that, in a state with wave function $\Psi_1(q)$, some measurement leads with certainty to a definite result (result 1), while in a state with $\Psi_2(q)$ it leads to result 2. Then it is asserted that every linear combination of Ψ_1 and Ψ_2, i.e. every function of the form $c_1\Psi_1+c_2\Psi_2$ (where c_1 and c_2 are constants), describes a state in which that measurement leads to either result 1 or result 2. Moreover, we can assert that, if we know the time dependence of the states, which for the one case is given by the function $\Psi_1(q, t)$, and for the other by $\Psi_2(q, t)$, then any linear combination also gives a possible dependence of a state on time.

These propositions constitute what is called the *principle of superposition of states*. In particular, it follows from this principle that equations satisfied by wave functions must be linear.

Let us consider a system composed of two parts, and suppose that the state of this system is given in such a way that each of its parts is completely described.[†] Then we can say that the probabilities of the coordinates q_1 of the first part are independent of the probabilities of the coordinates q_2 of the second part, and therefore the probability distribution for the whole system should be equal to the product of the probabilities of its parts. This means that the wave function $\Psi_{12}(q_1, q_2)$ of the system can be represented in the form of a product of the wave functions $\Psi_1(q_1)$ and $\Psi_2(q_2)$ of its parts:

$$\Psi_{12}(q_1, q_2) = \Psi_1(q_1)\, \Psi_2(q_2). \qquad (2.3)$$

If the two parts do not interact, then this relation between the wave function of the system and those of its parts will be maintained at future instants also:

$$\Psi_{12}(q_1, q_2, t) = \Psi_1(q_1, t)\, \Psi_2(q_2, t). \qquad (2.4)$$

§3. Operators

Let us consider some physical quantity f which characterises the state of a quantum system. Strictly, we should speak in the following discussion not of one quantity, but of a complete set of them at the

---

[†] This, of course, means that the state of the whole system is completely described also. However, we emphasise that the converse statement is by no means true: a complete description of the state of the whole system does not in general completely determine the states of its individual parts (we shall return to this point in §7).

same time. However, the discussion is not essentially changed by this, and for brevity and simplicity we shall work below in terms of only one physical quantity.

The values which a given physical quantity can take are called in quantum mechanics its *eigenvalues*, and the set of these is referred to as the *spectrum of eigenvalues* of the given quantity. In classical mechanics, generally speaking, quantities run through a continuous series of values. In quantum mechanics also there are physical quantities (for instance, the coordinates) whose eigenvalues occupy a continuous range; in such cases we speak of a *continuous spectrum* of eigenvalues. As well as such quantities, however, there exist in quantum mechanics others whose eigenvalues form some discrete set; in such cases we speak of a *discrete spectrum.*

We shall suppose for simplicity that the quantity f considered here has a discrete spectrum; the case of a continuous spectrum will be discussed in §5. The eigenvalues of the quantity f are denoted by f_n, where the suffix n takes the values 0, 1, 2, 3, ... We also denote the wave function of the system, in the state where the quantity f has the value f_n, by Ψ_n. The wave functions Ψ_n are called the *eigenfunctions* of the given physical quantity f. Each of these functions is supposed normalised, so that

$$\int |\Psi_n|^2 \, \mathrm{d}q = 1. \tag{3.1}$$

If the system is in some arbitrary state with wave function Ψ, a measurement of the quantity f carried out on it will give as a result one of the eigenvalues f_n. In accordance with the principle of superposition, we can assert that the wave function Ψ must be a linear combination of those eigenfunctions Ψ_n which correspond to the values f_n that can be obtained, with probability different from zero, when a measurement is made on the system and it is in the state considered. Hence, in the general case of an arbitrary state, the function Ψ can be represented in the form of a series

$$\Psi = \sum a_n \Psi_n, \tag{3.2}$$

where the summation extends over all n, and the a_n are some constant coefficients.

Thus we reach the conclusion that any wave function can be, as we say, expanded in terms of the eigenfunctions of any physical quantity. A set of functions in terms of which such an expansion can be made is called a *complete set*.

The expansion (3.2) makes it possible to determine the probability of finding (by measurement), in a system in a state with wave function Ψ, any given value f_n of the quantity f. For, according to what was said in the previous section, these probabilities must be determined by some expressions bilinear in Ψ and Ψ^*, and therefore must be bilinear in a_n and a_n^*. Furthermore, these expressions must, of course, be positive. Finally, the probability of the value f_n must become unity if the system is in a state with wave function $\Psi = \Psi_n$, and must become zero if there is no term containing Ψ_n in the expansion (3.2). The only essentially positive quantity satisfying this condition is the square of the modulus of the coefficient a_n. Thus we reach the result that the squared modulus $|a_n|^2$ of each coefficient in the expansion (3.2) determines the probability of the corresponding value f_n of the quantity f in the state with wave function Ψ. The sum of the probabilities of all possible values f_n must be equal to unity; in other words, the relation

$$\sum_n |a_n|^2 = 1 \tag{3.3}$$

must hold.

We shall now introduce the concept of the *mean value* $\bar{f}$ of the quantity f in the given state. In accordance with the usual definition of mean values, we define $\bar{f}$ as the sum of all the eigenvalues f_n of the given quantity, each multiplied by the corresponding probability $|a_n|^2$. Thus

$$\bar{f} = \sum_n f_n |a_n|^2 . \tag{3.4}$$

We shall write $\bar{f}$ in the form of an expression which does not contain the coefficients a_n in the expansion of the function Ψ, but this function itself. Since the products $a_n^* a_n$ appear in (3.4), it is clear that the required expression must be bilinear in Ψ^* and Ψ. We introduce a mathematical *operator*, which we denote[†] by $\hat{f}$ and define as follows. Let $(\hat{f}\Psi)$

[†] By convention, we shall always denote operators by letters with circumflexes.

denote the result of the operator $\hat{f}$ acting on the function Ψ. We define $\hat{f}$ in such a way that the integral of the product of $(\hat{f}\Psi)$ and the complex conjugate function Ψ^* is equal to the mean value $\bar{f}$:

$$\bar{f} = \int \Psi^*(\hat{f}\Psi)\,\mathrm{d}q. \tag{3.5}$$

Since the expression (3.5) is bilinear in Ψ^* and Ψ, the operator $\hat{f}$ itself must be what is called a *linear* operator. This term denotes operators having the properties†

$$\hat{f}(\Psi_1+\Psi_2) = \hat{f}\Psi_1+\hat{f}\Psi_2, \quad \hat{f}(a\Psi) = a\hat{f}\Psi,$$

where Ψ_1 and Ψ_2 are arbitrary functions and a is an arbitrary constant.

Thus, for every physical quantity in quantum mechanics, there is a definite corresponding linear operator.

If the function Ψ is one of the eigenfunctions Ψ_n, the mean value $\bar{f}$ must be equal to the definite value f_n that the quantity f has in the state concerned:

$$\bar{f} = \int \Psi_n^*\hat{f}\Psi_n\,\mathrm{d}q = f_n.$$

For this to be so, we must evidently have

$$\hat{f}\Psi_n = f_n\Psi_n, \tag{3.6}$$

so that the effect of the operator $\hat{f}$ on the eigenfunction Ψ_n is simply to multiply it by the corresponding eigenvalue f_n.

The eigenvalues of a physical quantity are therefore solutions of the equation

$$\hat{f}\Psi = f\Psi, \tag{3.7}$$

where f is a constant, and the eigenvalues are the values of this constant for which (3.7) has solutions satisfying the required conditions. As we shall see later, the form of the operators for various physical quantities can be established from straightforward physical arguments; this pro-

---

† Where no misunderstanding is possible, we shall usually omit the parentheses from the expression $(\hat{f}\Psi)$, the operator being understood to act on the expression that follows it.

perty of the operators then enables us to find the eigenfunctions and eigenvalues by solving the equation (3.7).

Both the eigenvalues of a real physical quantity and its mean value in every state must be real. This imposes a restriction on the corresponding operators. Equating the expression (3.5) to its complex conjugate, we obtain the relation

$$\int \Psi^*(\hat{f}\Psi)\,\mathrm{d}q = \int \Psi(\hat{f}^*\Psi^*)\,\mathrm{d}q, \tag{3.8}$$

where $\hat{f}^*$ denotes the operator which is the complex conjugate of $\hat{f}$. This relation does not hold in general for an arbitrary linear operator, so that it is a restriction on the form of the operator $\hat{f}$. For an arbitrary operator $\hat{f}$ we can find what is called the *transposed operator* $\tilde{\hat{f}}$, defined in such a way that

$$\int \Phi(\hat{f}\Psi)\,\mathrm{d}q = \int \Psi(\tilde{\hat{f}}\Phi)\,\mathrm{d}q, \tag{3.9}$$

where Ψ and Φ are two different functions. If we take, as the function Φ, the function Ψ^* which is the complex conjugate of Ψ, then a comparison with (3.8) shows that we must have

$$\tilde{\hat{f}} = \hat{f}^*. \tag{3.10}$$

Operators satisfying this condition are said to be *Hermitian*. Thus the operators corresponding, in the mathematical formalism of quantum mechanics, to real physical quantities must be Hermitian.

We can formally consider complex physical quantities also, i.e. those whose eigenvalues are complex. Let f be such a quantity. Then we can introduce its complex conjugate quantity f^*, whose eigenvalues are the complex conjugates of those of f. We denote by $\hat{f}^+$ the operator corresponding to the quantity f^*. It is called the *Hermitian conjugate* of the operator $\hat{f}$ and, in general, will be different from the complex conjugate operator $\hat{f}^*$: from the definition of the operator $\hat{f}^+$, the mean value of the quantity f^* in a state Ψ is

$$\overline{f^*} = \int \Psi^*\hat{f}^+\Psi\,\mathrm{d}q.$$

We also have

$$(\bar{f})^* = \left[\int \Psi^* f \hat{\Psi}\, \mathrm{d}q\right]^* = \int \Psi \hat{f}^* \Psi^*\, \mathrm{d}q = \int \Psi^* \tilde{\hat{f}} \Psi \mathrm{d}q.$$

Equating these two expressions gives

$$\hat{f}^+ = \tilde{\hat{f}}^*, \tag{3.11}$$

whence it is clear that $\hat{f}^+$ is in general not the same as $\hat{f}^*$. The condition (3.10) can now be written

$$\hat{f} = \hat{f}^+, \tag{3.12}$$

i.e. the operator of a real physical quantity is the same as its Hermitian conjugate; for this reason, Hermitian operators are also called *self-conjugate.*

Let f_n and f_m be two different eigenvalues of the real physical quantity f, and Ψ_n, Ψ_m the corresponding eigenfunctions:

$$\hat{f}\Psi_n = f_n \Psi_n, \quad \hat{f}\Psi_m = f_m \Psi_m.$$

Multiplying both sides of the first of these equations by Ψ_m^*, and both sides of the complex conjugate of the second by Ψ_n, and subtracting corresponding terms, we find

$$\Psi_m^* \hat{f} \Psi_n - \Psi_n \hat{f}^* \Psi_m^* = (f_n - f_m)\, \Psi_n \Psi_m^*.$$

We integrate both sides of this equation over q. Since $\hat{f}^* = \tilde{\hat{f}}$, by (3.9) the integral on the left-hand side of the equation is zero, so that we have

$$(f_n - f_m) \int \Psi_n \Psi_m^*\, \mathrm{d}q = 0.$$

If $f_n \neq f_m$, it follows that

$$\int \Psi_n \Psi_m^*\, \mathrm{d}q = 0,$$

and the different eigenfunctions are *orthogonal.* Together with the normalisation condition for these functions, the result may be written

$$\int \Psi_n \Psi_m^*\, \mathrm{d}q = \delta_{nm}, \tag{3.13}$$

where $\delta_{nm} = 1$ for $n = m$ and $\delta_{nm} = 0$ for $n \neq m$. Thus the eigenfunctions Ψ_n form a complete set of normalised and orthogonal (or, for brevity, *orthonormal*) functions.

The coefficients a_n in the expansion (3.2) are now easily determined. We need only multiply both sides of (3.2) by Ψ_m^* and integrate over q. By (3.13), all the terms in the sum are zero except the one with $n=m$, and the result is

$$a_m = \int \Psi\Psi_m^* \, dq. \tag{3.14}$$

We have spoken here of only one physical quantity f, whereas, as we said at the beginning of this section, we should have spoken of a complete set of simultaneously measurable physical quantities. We should then have found that to each of these quantities f, g,... there corresponds its operator $\hat{f}$, $\hat{g}$, ... The eigenfunctions Ψ_n then correspond to states in which all the quantities concerned have definite values, i.e. they correspond to definite sets of eigenvalues f_n, g_n, ..., and are simultaneous solutions of the system of equations

$$\hat{f}\Psi = f\Psi, \quad \hat{g}\Psi = g\Psi, \ldots.$$

§4. Addition and multiplication of operators

If $\hat{f}$ and $\hat{g}$ are the operators corresponding to two physical quantities f and g, the operator $\hat{f} + \hat{g}$ will correspond to the sum $f+g$. The significance of adding two different physical quantities in quantum mechanics depends considerably on whether the quantities can be measured simultaneously. If f and g are simultaneously measurable, the operators $\hat{f}$ and $\hat{g}$ have common eigenfunctions, which are also eigenfunctions of the operator $\hat{f} + \hat{g}$, and the eigenvalues of the latter are equal to the the sums $f_n + g_n$.

If, however, the quantities f and g cannot simultaneously take definite values, their sum $f+g$ has a more restricted significance. We can assert only that its mean value in an arbitrary state is equal to the sum of the mean values of the separate quantities:

$$\overline{f+g} = \bar{f} + \bar{g}. \tag{4.1}$$

The eigenvalues and eigenfunctions of the operator $\hat{f}+\hat{g}$ will not in general bear any relation to those of f and g. If $\hat{f}$ and $\hat{g}$ are Hermitian operators, then obviously $\hat{f}+\hat{g}$ is also Hermitian, and its eigenvalues are real and equal to those of the quantity $f+g$ defined in this way.

Next, let f and g once more be quantities that can be measured simultaneously. Besides their sum, we can also introduce the concept of their *product* as being a quantity whose eigenvalues are equal to the products of those of the quantities f and g. It is easy to see that, to this quantity, there corresponds an operator whose effect consists of the successive action on the function of first one and then the other operator. Such an operator is represented mathematically by the product of the operators $\hat{f}$ and $\hat{g}$. For, if Ψ_n are the eigenfunctions common to the operators $\hat{f}$ and $\hat{g}$, we have

$$\hat{f}\hat{g}\Psi_n = \hat{f}(\hat{g}\Psi_n) = \hat{f}g_n\Psi_n = g_n\hat{f}\Psi_n = g_nf_n\Psi_n \tag{4.2}$$

(the symbol $\hat{f}\hat{g}$ denotes an operator whose effect on a function Ψ consists of the successive action first of the operator $\hat{g}$ on the function Ψ and then of the operator $\hat{f}$ on the function $\hat{g}\Psi$). We could equally well take the operator $\hat{g}\hat{f}$ instead of $\hat{f}\hat{g}$, the former differing from the latter in the order of its factors. It is obvious that the result of the action of either of these operators on the functions Ψ_n will be the same. Since, however, every wave function Ψ can be represented as a linear combination of the functions Ψ_n, it follows that the result of the action of the operators $\hat{f}\hat{g}$ and $\hat{g}\hat{f}$ on an arbitrary function will also be the same. This fact can be written in the form of the symbolic equation $\hat{f}\hat{g} = \hat{g}\hat{f}$ or

$$\hat{f}\hat{g}-\hat{g}\hat{f} = 0. \tag{4.3}$$

Two such operators $\hat{f}$ and $\hat{g}$ are said to *commute* with each other.[†] Thus we arrive at the important result: if two quantities f and g can simultaneously take definite values, then their operators commute with each other.

The converse theorem can also be proved: if the operators $\hat{f}$ and $\hat{g}$

[†] The difference $\hat{f}\hat{g}-\hat{g}\hat{f}$ is called the *commutator* of the two operators.

commute, then all their eigenfunctions can be taken common to both; physically, this means that the corresponding physical quantities can be measured simultaneously. Thus the commutability of the operators is a necessary and sufficient condition for the physical quantities to be simultaneously measurable.

If the quantities f and g cannot be measured simultaneously, the concept of their product cannot be defined in the above manner. This appears in the fact that the operator $\hat{f}\hat{g}$ is not Hermitian in this case, and hence cannot correspond to any physical quantity. For, by the definition of the transpose of an operator we can write

$$\int \Psi\hat{f}\hat{g}\Phi \, dq = \int (\hat{g}\Phi)\left(\tilde{\hat{f}}\Psi\right) dq.$$

Here the operator $\tilde{\hat{f}}$ acts only on the function Ψ, and the operator $\hat{g}$ on Φ. Again using the definition of the transpose of an operator, we can write

$$\int \Psi\hat{f}\hat{g}\Phi \, dq = \int \left(\tilde{\hat{f}}\Psi\right)(\hat{g}\Phi) \, dq = \int \Phi\tilde{\hat{g}}\tilde{\hat{f}}\Psi \, dq.$$

Thus we obtain an integral in which the functions Ψ and Φ have changed places as compared with the original one. In other words, the operator $\tilde{\hat{g}}\tilde{\hat{f}}$ is the transpose of $\hat{f}\hat{g}$, and we can write

$$\widetilde{\hat{f}\hat{g}} = \tilde{\hat{g}}\tilde{\hat{f}}, \tag{4.4}$$

i.e. the transpose of the product $\hat{f}\hat{g}$ is the product of the transposes of the factors written in the opposite order. Taking the complex conjugate of both sides of equation (4.4), we have

$$(\hat{f}\hat{g})^{+} = \hat{g}^{+}\hat{f}^{+}. \tag{4.5}$$

If each of the operators $\hat{f}$ and $\hat{g}$ is Hermitian, then $(\hat{f}\hat{g})^{+} = \hat{g}\hat{f}$. It follows from this that the operator $\hat{f}\hat{g}$ is Hermitian if and only if the factors $\hat{f}$ and $\hat{g}$ commute.

§5. The continuous spectrum

All the relations given in §§3 and 4, describing the properties of the eigenfunctions of a discrete spectrum, can be generalised without difficulty to the case of a continuous spectrum of eigenvalues. We shall enumerate the results here without repeating the corresponding derivations.

Let f be a physical quantity having a continuous spectrum. We shall denote its eigenvalues by the same letter f simply, without suffix, and the corresponding eigenfunctions by Ψ_f. Just as an arbitrary wave function Ψ can be expanded in a series (3.2) of eigenfunctions of a quantity having a discrete spectrum, it can also be expanded (this time as an integral) in terms of the complete set of eigenfunctions of a quantity with a continuous spectrum. This expansion has the form

$$\Psi(q) = \int a_f \Psi_f(q)\,\mathrm{d}f. \tag{5.1}$$

The expansion coefficients are

$$a_f = \int \Psi(q)\,\Psi_f^*(q)\,\mathrm{d}q. \tag{5.2}$$

Since f can take a continuous range of values, we must now speak not of the probability of a particular value but of the probability of a value in an infinitesimal range between f and $f+\mathrm{d}f$. This probability is $|a_f|^2\,\mathrm{d}f$, just as $|a_n|^2$ gives the probability of the eigenvalue f_n for a discrete spectrum. Since the sum of the probabilities of all possible values of f must be equal to unity, we have

$$\int |a_f|^2\,\mathrm{d}f = 1 \tag{5.3}$$

(similarly to the relation (3.3) for a discrete spectrum).

The above formulae presuppose a particular normalisation of the eigenfunctions Ψ_f: they must be normalised according to

$$\int \Psi_{f'}^*\Psi_f\,\mathrm{d}q = \delta(f'-f), \tag{5.4}$$

the function on the right being a δ-function, whose definition and properties have been given in *Mechanics and Electrodynamics*, §54.[†]

[†] The δ-function was first used in theoretical physics by Dirac.

For, if (5.1) is substituted in (5.2), we get

$$a_f = \int a_{f'}\left(\int \Psi_{f'}\Psi_f^*\, \mathrm{d}q\right) \mathrm{d}f',$$

which must be satisfied identically. The condition (5.4) ensures that this is so, since the properties of the δ-function give

$$\int a_{f'}\, \delta(f'-f)\, \mathrm{d}f' = a_f.$$

The normalisation rule (5.4) replaces the condition (3.13) for the discrete spectrum. We see that the functions Ψ_f and $\Psi_{f'}$ with $f \neq f'$ are, as before, orthogonal. The integrals of the squares $|\Psi_f|^2$ of the eigenfunctions of a continuous spectrum are, however, infinite. The origin and significance of this divergence will be discussed at the end of §10.

Substitution of (5.2) in (5.1) gives

$$\Psi(q) = \int \Psi(q')\left(\int \Psi_f^*(q')\, \Psi_f(q)\, \mathrm{d}f\right) \mathrm{d}q',$$

whence it follows that[†]

$$\int \Psi_f^*(q')\, \Psi_f(q)\, \mathrm{d}f = \delta(q-q'). \tag{5.5}$$

Comparing the pair of formulae (5.1), (5.4) with the pair (5.2), (5.5), we see that, on the one hand, the functions $\Psi_f(q)$ provide an expansion of an arbitrary function $\Psi(q)$ with expansion coefficients a_f and, on the other hand, formula (5.2) represents an entirely analogous expansion of the function $a_f \equiv a(f)$ in terms of the functions $\Psi_f^*(q)$, while the $\Psi(q)$ play the part of expansion coefficients. The function $a(f)$, like $\Psi(q)$, completely determines the state of the system; it is called a wave function *in the f representation* (while the function $\Psi(q)$ is called a wave function in the coordinate or q representation). Just as $|\Psi(q)|^2$ determines the probability for the system to have coordinates lying in a given interval $\mathrm{d}q$, so $|a(f)|^2$ determines the probability for the values of the quantity f to lie in a given interval $\mathrm{d}f$.

[†] There is, of course, an analogous relation for a discrete spectrum:

$$\sum_n \Psi_n^*(q')\, \Psi_n(q) = \delta(q-q'). \tag{5.5a}$$

On the one hand, the functions $\Psi_f(q)$ are the eigenfunctions of the quantity f in the q representation; on the other hand, their complex conjugates are the eigenfunctions of the coordinate q in the f representation.

There are also physical quantities which in one range of values have a discrete spectrum, and in another a continuous spectrum. For the eigenfunctions of such a quantity all the relations derived in this and the previous sections are, of course, true. It need only be noted that the complete set of functions is formed by combining the eigenfunctions of both spectra. Hence the expansion of an arbitrary wave function in terms of the eigenfunctions of such a quantity has the form

$$\Psi(q) = \sum_n a_n\Psi_n(q) + \int a_f\Psi_f(q)\,\mathrm{d}f, \tag{5.6}$$

where the sum is taken over the discrete spectrum and the integral over the whole continuous spectrum.

The coordinate q itself is an example of a quantity having a continuous spectrum. It is easy to see that the operator corresponding to it is simply multiplication by q. For, since the probability of the various values of the coordinate is determined by the square $|\Psi(q)|^2$, the mean value of the coordinate is

$$\bar{q} = \int q\,|\Psi|^2\,\mathrm{d}q \equiv \int \Psi^* q \Psi\,\mathrm{d}q.$$

Comparison of this expression with the definition (3.5) of the operator $\hat{f}$ shows that[†]

$$\hat{q} = q. \tag{5.7}$$

The eigenfunctions of this operator must be determined, according to the usual rule, by the equation $q\Psi_{q_0} = q_0\Psi_{q_0}$, where q_0 temporarily denotes the actual values of the coordinate as distinct from the variable q. Since this equation can be satisfied either by $\Psi_{q_0} = 0$ or by $q = q_0$, it is clear that the wave functions which satisfy the normalisation condition are

$$\Psi_{q_0} = \delta(q-q_0). \tag{5.8}$$

[†] In future we shall always, for simplicity, write operators which amount to multiplication by some quantity in the form of that quantity itself.

§6. The passage to the limiting case of classical mechanics

Quantum mechanics contains classical mechanics in the form of a certain limiting case. The question arises as to how this passage to the limit is made.

In quantum mechanics an electron is described by a wave function which determines the various values of its coordinates; of this function we so far know only that it is the solution of a certain linear partial differential equation. In classical mechanics, on the other hand, an electron is regarded as a material particle, moving in a path which is completely determined by the equations of motion. There is an interrelation, somewhat similar to that between quantum and classical mechanics, in electrodynamics between wave optics and geometrical optics. In wave optics, the electromagnetic waves are described by the electric and magnetic field vectors, which satisfy a definite system of linear differential equations, namely Maxwell's equations. In geometrical optics, however, the propagation of light along definite paths, or rays, is considered. Such an analogy enables us to see that the passage from quantum mechanics to the limit of classical mechanics occurs similarly to the passage from wave optics to geometrical optics.

Let us recall how this latter transition is made mathematically (see *Mechanics and Electrodynamics*, §74). Let u be any of the field components in the electromagnetic wave. It can be written in the form $u = ae^{i\phi}$ (with a and ϕ real), where a is called the *amplitude* and ϕ the *phase* of the wave (called in geometrical optics the *eikonal*). The limiting case of geometrical optics corresponds to small wavelengths; this is expressed mathematically by saying that ϕ varies by a large amount over short distances; this means, in particular, that it can be supposed large in absolute value.

Similarly, we start from the hypothesis that, to the limiting case of classical mechanics, there correspond in quantum mechanics wave functions of the form $\Psi = ae^{i\phi}$, where a is a slowly varying function and ϕ takes large values. As is well known, the path of a particle can be determined in mechanics by means of the variational principle, according to which what is called the *action* S of a mechanical system must take its least possible value (the *principle of least action*). In geometrical optics the path of the rays is determined by what is called *Fermat's principle*, according to which the *optical path length* of the ray, i.e. the difference between its phases at the beginning and end of the path, must take its least (or greatest) possible value.

On the basis of this analogy, we can assert that the phase of the wave function, in the limiting (classical) case, must be proportional to the mechanical action S of the physical system considered, i.e. we must have $S = \text{constant} \times \phi$. The constant of proportionality is called *Planck's constant*[†] and is denoted by $\hbar$. It has the dimensions of action (since ϕ is dimensionless) and has the value

$$\hbar = 1.054 \times 10^{-27} \text{ erg sec.}$$

Thus, the wave function of an "almost classical" (or, as we say, *quasi-classical*) physical system has the form

$$\Psi = ae^{iS/\hbar}. \tag{6.1}$$

Planck's constant $\hbar$ plays a fundamental part in all quantum phenomena. Its relative value (compared with other quantities of the same dimensions) determines the "extent of quantisation" of a given physical system. The transition from quantum mechanics to classical mechanics, corresponding to large phase, can be formally described as a passage to the limit $\hbar \to 0$ (just as the transition from wave optics to geometrical optics corresponds to a passage to the limit of zero wavelength, $\lambda \to 0$).

We have ascertained the limiting form of the wave function, but the question still remains how it is related to classical motion in a path. In general, the motion described by the wave function does not tend to motion in a definite path. Its connection with classical motion is that, if at some initial instant the wave function, and with it the probability distribution of the coordinates, is given, then at subsequent

---

[†] It was introduced into physics by M. Planck in 1900. The constant $\hbar$, which we use everywhere in this book, is, strictly speaking, Planck's constant divided by 2π; this is Dirac's notation.

instants this distribution will change according to the laws of classical mechanics (for a more detailed discussion of this, see §26).

In order to obtain motion in a definite path, we must start from a wave function of a particular form, which is perceptibly different from zero only in a very small region of space (what is called a *wave packet*); the dimensions of this region must tend to zero with $\hbar$. Then we can say that, in the quasi-classical case, the wave packet will move in space along a classical path of a particle.

Finally, quantum-mechanical operators must reduce, in the limit, simply to multiplication by the corresponding physical quantity.

§7. The density matrix

The description of a system by means of a wave function is the most complete description possible in quantum mechanics, in the sense indicated at the end of §1.

States that do not allow such a description are encountered if we consider a system that is part of a larger closed system.

We suppose that the closed system as a whole is in some state described by the wave function $\Psi(q, x)$, where x denotes the set of coordinates of the system considered, and q the remaining coordinates of the closed system. This function is general does not fall into a product of functions of x and of q alone, so that the system does not have its own wave function.

Let f be some physical quantity pertaining to the system considered. Its operator therefore acts only on the coordinates x, and not on q. The mean value of this quantity in the state considered is

$$\bar{f} = \iint \Psi^*(q, x) \hat{f} \Psi(q, x) \, \mathrm{d}q \, \mathrm{d}x. \tag{7.1}$$

We introduce the function $\varrho(x', x)$ defined by

$$\varrho(x', x) = \int \Psi^*(q, x') \, \Psi(q, x) \, \mathrm{d}q, \tag{7.2}$$

where the integration is extended only over the coordinates q; this function is called the *density matrix* of the system. When $x = x'$, we have the function

$$\varrho(x, x) = \int |\Psi^*(q, x)|^2 \, dq, \tag{7.3}$$

which evidently determines the probability distribution for the coordinates of the system.

Using the density matrix, the mean value $\bar{f}$ can be written in the form

$$\bar{f} = \int [\hat{f}\varrho(x', x)]_{x'=x} \, dx. \tag{7.4}$$

Here $\hat{f}$ acts only on the variables x in the function $\varrho(x', x)$; after calculating the result of its action, we put $x' = x$. We see that, if we know the density matrix, we can calculate the mean value of any quantity characterising the system. It follows from this that, by means of $\varrho(x', x)$, we can also determine the probabilities of various values of the physical quantities in the system. Thus we reach the conclusion that the state of a system which does not have a wave function can be described by means of a density matrix.† This does not contain the coordinates q which do not belong to the system concerned, though, of course, it depends essentially on the state of the closed system as a whole.

The description by means of the density matrix is the most general form of quantum-mechanical description of the system. The description by means of the wave function, on the other hand, is a particular case of this, corresponding to a density matrix of the form $\varrho(x', x) = \Psi^*(x')\Psi(x)$. The following important difference exists between this particular case and the general one. For a state having a wave function (sometimes called a *pure state*) there is always a complete set of measuring processes such that they lead with certainty to definite results. For states having only a density matrix (called *mixed states*), on the other hand, there is no complete set of measuring processes whose result can be uniquely predicted.

---

† The quantum-mechanical description of such states was introduced independently by L. D. Landau and F. Bloch in 1927.

CHAPTER 2

CONSERVATION LAWS IN QUANTUM MECHANICS

§8. The Hamiltonian operator

The wave function Ψ completely determines the state of a physical system in quantum mechanics. This means that, if this function is given at some instant, not only are all the properties of the system at that instant described, but its behaviour at all subsequent instants is determined (only, of course, to the degree of completeness which is generally admissible in quantum mechanics). The mathematical expression of this fact is that the value of the derivative $\partial\Psi/\partial t$ of the wave function with respect to time at any given instant must be determined by the value of the function itself at that instant, and, by the principle of superposition, the relation between them must be linear. In the most general form we can write

$$i\hbar\, \partial\Psi/\partial t = \hat{H}\Psi, \tag{8.1}$$

where $\hat{H}$ is some linear operator; the reason for the factor $i\hbar$ will be explained later.

Since the integral $\int \Psi^*\Psi\, \mathrm{d}q$ is a constant independent of time, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \Psi^*\Psi\, \mathrm{d}q = \int \Psi^* \frac{\partial\Psi}{\partial t}\, \mathrm{d}q + \int \frac{\partial\Psi^*}{\partial t}\, \Psi\, \mathrm{d}q = 0.$$

Substituting here from (8.1) and using in the second integral the definition of the transpose of an operator, we can write (omitting the common factor $1/i\hbar$)

$$\int \Psi^*\hat{H}\Psi\,\mathrm{d}q-\int \Psi\hat{H}^*\Psi^*\,\mathrm{d}q = \int \Psi^*\hat{H}\Psi\,\mathrm{d}q-\int \Psi^*\tilde{\hat{H}}^*\Psi\,\mathrm{d}q$$
$$= \int \Psi^*(\hat{H}-\hat{H}^+)\Psi\,\mathrm{d}q = 0.$$

Since this equation must hold for an arbitrary function Ψ, it follows that we must have identically $\hat{H} = \hat{H}^+$; the operator $\hat{H}$ is therefore Hermitian. Let us find the classical quantity to which the operator $\hat{H}$ corresponds. To do this, we use the limiting expression (6.1) for the wave function and write

$$\frac{\partial\Psi}{\partial t} = \frac{i}{\hbar}\frac{\partial S}{\partial t}\Psi;$$

the slowly varying amplitude a need not be differentiated. Comparing this equation with the definition (8.1), we see that, in the limiting case, the operator $\hat{H}$ reduces to simply multiplying by $-\partial S/\partial t$. This means that $-\partial S/\partial t$ is the physical quantity into which the Hermitian operator $\hat{H}$ passes.

The derivative $-\partial S/\partial t$ is just Hamilton's function H for a mechanical system. Thus the operator $\hat{H}$ is what corresponds in quantum mechanics to Hamilton's function; this operator is called the *Hamiltonian operator* or, more briefly, the *Hamiltonian* of the system. If the form of the Hamiltonian is known, equation (8.1) determines the wave function of the physical system concerned. This fundamental equation of quantum mechanics is called the *wave equation.*

§9. The differentiation of operators with respect to time

The concept of the derivative of a physical quantity with respect to time cannot be defined in quantum mechanics in the same way as in classical mechanics. For the definition of the derivative in classical mechanics involves the consideration of the values of the quantity at two neighbouring but distinct instants of time. In quantum mechanics, however, a quantity which at some instant is measured does not in general have definite values at subsequent instants; this was discussed in detail in §1.

Hence the idea of the derivative with respect to time must be differently defined in quantum mechanics. It is natural to define the *derivative* $\dot{f}$ of a quantity f as the quantity whose mean value is equal to the derivative, with respect to time, of the mean value $\bar{f}$. Thus we have by definition

$$\overline{\dot{f}} = \dot{\bar{f}}. \tag{9.1}$$

Starting from this definition, it is easy to obtain an expression for the quantum-mechanical operator $\hat{\dot{f}}$ corresponding to the quantity $\dot{f}$. We can write

$$\begin{aligned}\overline{\dot{f}} = \dot{\bar{f}} &= \frac{\mathrm{d}}{\mathrm{d}t}\int \Psi^* \hat{f}\Psi\,\mathrm{d}q \\ &= \int \Psi^* \frac{\partial \hat{f}}{\partial t}\Psi\,\mathrm{d}q + \int \frac{\partial \Psi^*}{\partial t}\hat{f}\Psi\,\mathrm{d}q + \int \Psi^* \hat{f}\frac{\partial \Psi}{\partial t}\,\mathrm{d}q.\end{aligned}$$

Here $\partial\hat{f}/\partial t$ is the operator obtained by differentiating the operator $\hat{f}$ with respect to time; $\hat{f}$ may depend on the time as a parameter. Substituting for $\partial\Psi/\partial t$, $\partial\Psi^*/\partial t$ their expressions according to (8.1), we obtain

$$\overline{\dot{f}} = \int \Psi^* \frac{\partial \hat{f}}{\partial t}\Psi\,\mathrm{d}q + \frac{i}{\hbar}\int (\hat{H}^*\Psi^*)\hat{f}\Psi\,\mathrm{d}q - \frac{i}{\hbar}\int \Psi^*\hat{f}(\hat{H}\Psi)\,\mathrm{d}q.$$

Since the operator $\hat{H}$ is Hermitian, we have

$$\int (\hat{H}^*\Psi^*)(\hat{f}\Psi)\,\mathrm{d}q = \int \Psi^*\hat{H}\hat{f}\Psi\,\mathrm{d}q:$$

thus

$$\overline{\dot{f}} = \int \Psi^*\left(\frac{\partial \hat{f}}{\partial t} + \frac{i}{\hbar}\hat{H}\hat{f} - \frac{i}{\hbar}\hat{f}\hat{H}\right)\Psi\,\mathrm{d}q.$$

Since, on the other hand, we must have, by the definition of mean values, $\overline{\dot{f}} = \int \Psi^*\hat{\dot{f}}\Psi\,\mathrm{d}q$, it is seen that the expression in parentheses under the integral is the required operator $\hat{\dot{f}}$:

$$\hat{\dot{f}} = \frac{\partial \hat{f}}{\partial t} + \frac{i}{\hbar}(\hat{H}\hat{f} - \hat{f}\hat{H}). \tag{9.2}$$

If the operator $\hat{f}$ does not depend explicitly on time, $\hat{\dot{f}}$ reduces, apart from a constant factor, to the commutator of the operator $\hat{f}$ and the Hamiltonian.

A very important class of physical quantities is formed by those whose operators do not depend explicitly on time, and also commute with the Hamiltonian, so that $\hat{\dot{f}} = 0$. Such quantities are said to be *conserved*. For these, $\overline{\dot{f}} = \dot{\bar{f}} = 0$, that is, $\bar{f}$ is constant. In other words, the mean value of the quantity remains constant in time. We can also assert that, if in a given state the quantity f has a definite value (i.e. the wave function is an eigenfunction of the operator $\hat{f}$), then it will have a definite value (the same one) at subsequent instants also.

§10. Stationary states

If the system is closed or is in a *constant* external field, its Hamiltonian cannot contain the time explicitly. This follows from the fact that all times are equivalent so far as the given physical system is concerned. Since, on the other hand, any operator of course commutes with itself, we reach the conclusion that Hamilton's function is conserved for systems which are not in a varying external field. As is well known, a Hamilton's function which is conserved is called the *energy* (see *Mechanics and Electrodynamics*, §6). The law of conservation of energy in quantum mechanics signifies that, if in a given state the energy has a definite value, this value remains constant in time.

States in which the energy has definite values are called *stationary states* of a system. They are described by wave functions Ψ_n which are the eigenfunctions of the Hamiltonian operator, i.e. which satisfy the equation $\hat{H}\Psi_n = E_n\Psi_n$, where E_n are the eigenvalues of the energy. Correspondingly, the wave equation (8.1) for the function Ψ_n,

$$i\hbar\, \partial\Psi_n/\partial t = \hat{H}\Psi_n = E_n\Psi_n,$$

can be integrated at once with respect to time and gives

$$\Psi_n = e^{-(i/\hbar)E_n t}\psi_n(q), \tag{10.1}$$

where ψ_n is a function of the coordinates only. This determines the relation between the wave functions of stationary states and the time.

We shall denote by the small letter ψ the wave functions of stationary states without the time factor. These functions, and also the eigenvalues of the energy, are determined by the equation

$$\hat{H}\psi = E\psi. \tag{10.2}$$

The stationary state with the smallest possible value of the energy is called the *normal* or *ground state* of the system.

The expansion of an arbitrary wave function Ψ in terms of the wave functions of stationary states has the form

$$\Psi = \sum_n a_n e^{-(i/\hbar)E_n t}\psi_n(q). \tag{10.3}$$

The squared moduli $|a_n|^2$ of the expansion coefficients, as usual, determine the probabilities of various values of the energy of the system.

The probability distribution for the coordinates in a stationary state is determined by the squared modulus $|\Psi_n|^2 = |\psi_n|^2$; it is independent of time. The same is true of the mean values

$$\bar{f} = \int \Psi_n^* \hat{f} \Psi_n \, dq = \int \psi_n^* \hat{f} \psi_n \, dq$$

of any physical quantity f (whose operator does not depend explicitly on the time).

As has been said, the operator of any quantity that is conserved commutes with the Hamiltonian. This means that any physical quantity that is conserved can be measured simultaneously with the energy.

Among the various stationary states, there may be some which correspond to the same value of the energy (or, as we say, *energy level* of the system), but differ in the values of some other physical quantities. Such energy levels, to which several different stationary states correspond, are said to be *degenerate*. Physically, the possibility that degenerate levels can exist is related to the fact that the energy does not in general form by itself a complete set of physical quantities.

In particular, if there are two conserved physical quantities f and

g whose operators do not commute, then the energy levels of the system are in general degenerate. For, let ψ be the wave function of a stationary state in which, besides the energy, the quantity f also has a definite value. Then we can say that the function $\hat{g}\psi$ does not coincide (apart from a constant factor) with ψ; if it did, this would mean that the quantity g also had a definite value, which is impossible, since f and g cannot be measured simultaneously. On the other hand, the function $\hat{g}\psi$ is an eigenfunction of the Hamiltonian, corresponding to the same value E of the energy as ψ:

$$\hat{H}(\hat{g}\psi) = \hat{g}\hat{H}\psi = E(\hat{g}\psi).$$

Thus we see that the energy E corresponds to more than one eigenfunction, i.e. the energy level is degenerate.

It is clear that any linear combination of wave functions corresponding to the same degenerate energy level is also an eigenfunction for that value of the energy. In other words, the choice of eigenfunctions of a degenerate energy level is not unique. Arbitrarily selected eigenfunctions of a degenerate energy level are not, in general, orthogonal. By a proper choice of linear combinations of them, however, we can always obtain a set of orthogonal (and normalised) eigenfunctions.

The spectrum of eigenvalues of the energy may be either discrete or continuous. A stationary state of a discrete spectrum always corresponds to a *finite* motion of the system, i.e. one in which neither the system nor any part of it moves off to infinity. For, with eigenfunctions of a discrete spectrum, the integral $\int |\Psi|^2 \mathrm{d}q$, taken over all space, is finite. This certainly means that the squared modulus $|\Psi|^2$ decreases quite rapidly, becoming zero at infinity. In other words, the probability of infinite values of the coordinates is zero; that is, the system executes a finite motion, and is said to be in a *bound state*.

For wave functions of a continuous spectrum, the integral $\int |\Psi|^2 \mathrm{d}q$ diverges. Here the squared modulus $|\Psi|^2$ of the wave function does not directly determine the probability of the various values of the coordinates, and must be regarded only as a quantity proportional to this probability. The divergence of the integral $\int |\Psi|^2 \mathrm{d}q$ is always due to the fact that $|\Psi|^2$ does not become zero at infinity (or becomes zero insufficiently rapidly). Hence we can say that the integral $\int |\Psi|^2 \,\mathrm{d}q$, taken over the region of space outside any arbitrarily large but finite closed surface, will always diverge. This means that, in the state considered, the system (or some part of it) is at infinity. Thus the stationary states of a continuous spectrum correspond to an infinite motion of the system.

§11. Matrices of physical quantities

We shall suppose for convenience that the system considered has a discrete energy spectrum; all the relations obtained below can be generalised at once to the case of a continuous spectrum. Let $\Psi = \Sigma a_n \Psi_n$ be the expansion of an arbitrary wave function in terms of the wave functions of the stationary states. If we substitute this expansion in the definition (3.5) of the mean value of some quantity f, we obtain

$$\bar{f} = \sum_n \sum_m a_n^* a_m f_{nm}(t), \tag{11.1}$$

where $f_{nm}(t)$ denotes the integral

$$f_{nm} = \int \Psi_n^* \hat{f} \Psi_m \,\mathrm{d}q. \tag{11.2}$$

The set of quantities $f_{nm}(t)$ with all possible n and m is called the *matrix* of the quantity f, and each of the $f_{nm}(t)$ is called the *matrix element* corresponding to the *transition* from state m to state n.[†]

Another notation used for the matrix elements f_{nm} is

$$\langle n|f|m\rangle, \tag{11.3}$$

which is especially convenient when each of the suffixes n and m has to be written as a group of letters. The symbol (11.3) is sometimes regarded as being made up of f and the symbols $|m\rangle$ and $\langle n|$, which denote the initial and final states respectively; this is Dirac's notation.

[†] The matrix representation of physical quantities was introduced by Heisenberg in 1925, before Schrödinger's discovery of the wave equation. "Matrix mechanics" was later developed by M. Born, W. Heisenberg and P. Jordan.

The dependence of the matrix elements $f_{nm}(t)$ on time is determined (if the operator $\hat{f}$ does not contain the time explicitly) by the dependence of the functions Ψ_n on time. Substituting for them the expressions (10.1), we find that

$$f_{nm}(t) = f_{nm}\, e^{i\omega_{nm}t}, \tag{11.4}$$

where

$$\omega_{nm} = (E_n - E_m)/\hbar \tag{11.5}$$

is what is called the *transition frequency* between the states m and n, and the quantities

$$f_{nm} = \int \psi_n^* \hat{f} \psi_m \,\mathrm{d}q \tag{11.6}$$

form the matrix of the quantity f which is independent of time, and which is commonly used.

The matrix elements of the derivative $\dot{f}$ are obtained by differentiating the matrix elements of the quantity f with respect to time; this follows directly from the fact that the mean value $\overline{\dot{f}}$ is

$$\overline{\dot{f}} = \dot{\overline{f}} = \sum_n \sum_m a_n^* a_m \dot{f}_{nm}(t).$$

From (11.4) we thus have for the matrix elements of $\dot{f}$

$$\dot{f}_{nm}(t) = i\omega_{nm} f_{nm}(t) \tag{11.7}$$

or (cancelling the time factor $e^{i\omega_{nm}t}$ from both sides) for the matrix elements independent of time

$$(\dot{f})_{nm} = i\omega_{nm} f_{nm} = (i/\hbar)(E_n - E_m) f_{nm}. \tag{11.8}$$

To simplify the notation in the formulae, we shall derive all our relations below for the matrix elements independent of time; exactly similar relations hold for the matrices which depend on the time.

For the matrix elements of the complex conjugate f^* of the quantity f we obtain, taking into account the definition of the Hermitian conjugate operator,

$$(f^*)_{nm} = \int \psi_n^* \hat{f}^{+} \psi_m \,\mathrm{d}q = \int \psi_n^* \tilde{\hat{f}}^{*} \psi_m \,\mathrm{d}q = \int \psi_m \hat{f}^{*} \psi_n^* \,\mathrm{d}q$$

or

$$(f^*)_{nm} = (f_{mn})^*. \tag{11.9}$$

For real physical quantities, which are the only ones we usually consider, we consequently have

$$f_{nm} = f^*_{mn} \tag{11.10}$$

(f^*_{mn} stands for $(f_{mn})^*$). Such matrices, like the corresponding operators, are said to be *Hermitian*.

Matrix elements with $n = m$ are called *diagonal elements*. These are independent of time, and (11.10) shows that they are real. The element f_{nn} is the mean value of the quantity f in the state ψ_n.

It is not difficult to obtain the multiplication rule for matrices. To do so, we first observe that the formula

$$\hat{f}\psi_n = \sum_m f_{mn}\psi_m \tag{11.11}$$

holds. This is simply the expansion of the function $\hat{f}\psi_n$ in terms of the functions ψ_m, the coefficients being determined in accordance with the general formula (3.14). Remembering this formula, let us write down the result of the product of two operators acting on the function ψ_n:

$$\hat{f}\hat{g}\psi_n = \hat{f}\sum_k g_{kn}\psi_k = \sum_k g_{kn}\hat{f}\psi_k = \sum_{k,m} g_{kn}f_{mk}\psi_m.$$

Since, on the other hand, we must have

$$\hat{f}\hat{g}\psi_n = \sum_m (fg)_{mn}\psi_m,$$

we arrive at the result that the matrix elements of the product fg are determined by the formula

$$(fg)_{mn} = \sum_k f_{mk}g_{kn}. \tag{11.12}$$

This rule is the same as that used in mathematics for the multiplication of matrices: the rows of the first matrix in the product are multiplied by the columns of the second.

If the matrix is given, then so is the operator itself. In particular, if the matrix is given, it is in principle possible to determine the eigenvalues of the physical quantity concerned and the corresponding eigenfunctions.

We shall now consider the values of all quantities at some definite instant, and expand an arbitrary wave function Ψ (at that instant) in terms of the eigenfunctions of the Hamiltonian operator $\hat{H}$, i.e. of the wave functions ψ_m of the stationary states (these wave functions are independent of time):

$$\Psi = \sum_m c_m \psi_m, \tag{11.13}$$

where the expansion coefficients are denoted by c_m. We substitute this expansion in the equation $\hat{f}\Psi = f\Psi$ which determines the eigenvalues and eigenfunctions of the quantity f. We have

$$\sum_m c_m(\hat{f}\psi_m) = f \sum_m c_m \psi_m .$$

We multiply both sides of this equation by ψ_n^* and integrate over q. Each of the integrals $\int \psi_n^* \hat{f} \psi_m \, dq$ on the left-hand side of the equation is the corresponding matrix element f_{nm}. On the right-hand side, all the integrals $\int \psi_n^* \psi_m \, dq$ with $m \neq n$ vanish by virtue of the orthogonality of the functions ψ_m, and $\int \psi_n^* \psi_n \, dq = 1$ by virtue of their normalisation. Thus

$$\sum_m f_{nm} c_m = f c_n$$

or

$$\sum_m (f_{nm} - f\delta_{nm}) c_m = 0. \tag{11.14}$$

Thus we have obtained a system of homogeneous algebraic equations of the first degree (with the c_m as unknowns). As is well known, such a system has solutions different from zero only if the determinant formed by the coefficients in the equations vanishes, i.e. only if

$$|f_{nm} - f\delta_{nm}| = 0.$$

The roots of this equation (in which f is regarded as the unknown) are the possible values of the quantity f. The set of values c_m satisfying the equations (11.14) when f is equal to any of these values determines the corresponding eigenfunction.

If, in the definition (11.6) of the matrix elements of the quantity f, we take as ψ_n the eigenfunctions of this quantity, then from the equation $\hat{f}\psi_n = f_n\psi_n$ we have

$$f_{nm} = \int \psi_n^* \hat{f} \psi_m \,\mathrm{d}q = f_m \int \psi_n^* \psi_m \,\mathrm{d}q.$$

By virtue of the orthogonality and normalisation of the functions ψ_m, this gives $f_{nm} = 0$ for $n \neq m$ and $f_{mm} = f_m$. Thus only the diagonal matrix elements are different from zero, and each of these is equal to the corresponding eigenvalue of the quantity f. A matrix with only these elements different from zero is said to be put in *diagonal form.* In particular, in the usual representation, with the wave functions of the stationary states as the functions ψ_m, the energy matrix is diagonal (and so are the matrices of all other physical quantities having definite values in the stationary states). In general, the matrix of a quantity f, defined with respect to the eigenfunctions of some operator $\hat{g}$, is said to be the matrix *of f in a representation in which g is diagonal.* We shall always, except where the subject is specially mentioned, understand in future by the matrix of a physical quantity its matrix in the usual representation, in which the energy is diagonal. Everything that has been said above regarding the dependence of matrices on time refers, of course, only to this usual representation.†

§12. **Momentum**

Let us consider a closed system of particles. Since all positions in space of such a system as a whole are equivalent, we can say, in particular, that the Hamiltonian of the system does not vary when the system undergoes a parallel displacement over any distance. It is sufficient that this condition should be fulfilled for an arbitrary small displacement.

---

† Bearing in mind the diagonality of the energy matrix, it is easy to see that equation (11.8) is the operator relation (9.2) written in matrix form.

An infinitely small parallel displacement over a distance $\delta\mathbf{r}$ signifies a transformation under which the radius vectors $\mathbf{r}_a$ of all the particles (a being the number of the particle) receive the same increment $\delta\mathbf{r} : \mathbf{r}_a \rightarrow \mathbf{r}_a+\delta\mathbf{r}$. An arbitrary function $\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots)$ of the coordinates of the particles, under such a transformation, becomes the function

$$\begin{aligned}\psi(\mathbf{r}_1+\delta\mathbf{r}, \mathbf{r}_2+\delta\mathbf{r}, \ldots) &= \psi(\mathbf{r}_1, \mathbf{r}_2, \ldots)+\delta\mathbf{r}\,.\sum_a \nabla_a\psi \\ &= \left(1+\delta\mathbf{r}\,.\sum_a \nabla_a\right)\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots)\end{aligned}$$

(∇_a denotes the operator of differentiation with respect to $\mathbf{r}_a$). The expression

$$1+\delta\mathbf{r}\,.\sum_a \nabla_a \tag{12.1}$$

can be regarded as the operator of an infinitely small displacement, which converts the function $\psi(\mathbf{r}_1, \mathbf{r}_2, ...)$ into the function $\psi(\mathbf{r}_1+\delta\mathbf{r}, \mathbf{r}_2+\delta\mathbf{r}, \ldots)$.

The statement that some transformation does not change the Hamiltonian means that, if we make this transformation on the function $\hat{H}\psi$, the result is the same as if we make it only on the function ψ and then apply the operator $\hat{H}$. Mathematically, this can be written as follows. Let $\hat{O}$ be the operator which effects the transformation in question. Then we have $\hat{O}(\hat{H}\psi) = \hat{H}(\hat{O}\psi)$, whence

$$\hat{O}\hat{H}-\hat{H}\hat{O} = 0, \tag{12.2}$$

i.e. the Hamiltonian must commute with the operator $\hat{O}$.

In the case considered, the operator $\hat{O}$ is the operator (12.1) of an infinitely small displacement. Since the unit operator (the operator of multiplying by unity) commutes, of course, with any operator, and the constant factor $\delta\mathbf{r}$ can be taken in front of $\hat{H}$, the condition (12.2) reduces here to

$$\left(\sum_a \nabla_a\right)\hat{H}-\hat{H}\left(\sum_a \nabla_a\right) = 0. \tag{12.3}$$

As we know, the commutability of an operator (not containing the time explicitly) with $\hat{H}$ means that the physical quantity corresponding to that operator is conserved. The quantity whose conservation for a closed system follows from the homogeneity of space is the *momentum* of the system (see *Mechanics and Electrodynamics*, §7). Thus the relation (12.3) expresses the law of conservation of momentum in quantum mechanics; the operator $\Sigma\nabla_a$ must correspond, apart from a constant factor, to the total momentum of the system, and each term ∇_a of the sum to the momentum of an individual particle.

The coefficient of proportionality between the operator $\hat{\mathbf{p}}$ of the momentum of a particle and the operator ∇ can be determined by means of the passage to the limit of classical mechanics. Putting $\hat{\mathbf{p}} = c\triangle$ and using the limiting expression (6.1) for the wave function, we have

$$\hat{\mathbf{p}}\Psi = (i/\hbar)cae^{(i/\hbar)S}\,\nabla S = c(i/\hbar)\Psi\,\nabla S,$$

i.e. in the classical approximation the effect of the operator $\hat{\mathbf{p}}$ reduces to multiplication by $(i/\hbar)c\,\nabla S$. The gradient ∇S is the momentum $\mathbf{p}$ of the particle (see *Mechanics and Electrodynamics*, §31); hence we must have $c = -i\hbar$.

Thus the *momentum operator* of a particle is $\hat{\mathbf{p}} = -i\hbar\nabla$, or, in components,

$$\hat{p}_x = -i\hbar\,\partial/\partial x, \quad \hat{p}_y = -i\hbar\,\partial/\partial y, \quad \hat{p}_z = -i\hbar\,\partial/\partial z. \tag{12.4}$$

It is easy to see that these operators are Hermitian, as they should be. For, with arbitrary functions $\psi(x)$ and $\phi(x)$ which vanish at infinity, we have

$$\int \phi\hat{p}_x\psi\,\mathrm{d}x = -i\hbar\int \phi\frac{\partial\psi}{\partial x}\,\mathrm{d}x = i\hbar\int \psi\frac{\partial\phi}{\partial x}\,\mathrm{d}x = \int \psi\hat{p}_x^*\phi\,\mathrm{d}x,$$

and this is the condition that the operator should be Hermitian.

Since the result of differentiating functions with respect to two different variables is independent of the order of differentiation, it is clear that the operators of the three components of momentum commute with one another:

$$\hat{p}_x\hat{p}_y - \hat{p}_y\hat{p}_x = 0, \quad \hat{p}_x\hat{p}_z - \hat{p}_z\hat{p}_x = 0, \quad \hat{p}_y\hat{p}_z - \hat{p}_z\hat{p}_y = 0. \tag{12.5}$$

This means that all three components of the momentum of a particle can simultaneously have definite values.

Let us find the eigenfunctions and eigenvalues of the momentum operators. They are determined by the vector equation

$$-i\hbar\,\nabla\psi = \mathbf{p}\psi. \tag{12.6}$$

The solutions are of the form

$$\psi = Ce^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}}, \tag{12.7}$$

where C is a constant. If all three components of the momentum are given simultaneously, we see that this completely determines the wave function of the particle. In other words, the quantities p_x, p_y, p_z form one of the possible complete sets of physical quantities. Their eigenvalues form a continuous spectrum extending from $-\infty$ to $+\infty$.

According to the rule (5.4) for normalising the eigenfunctions of a continuous spectrum, we must have

$$\int \psi_{\mathbf{p}'}\psi_{\mathbf{p}}^*\,\mathrm{d}V = \delta(\mathbf{p}'-\mathbf{p}) \tag{12.8}$$

(where $\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$), the integration being extended over all space; $\delta(\mathbf{p}'-\mathbf{p})$ is the three-dimensional δ-function[†]. The integration can be effected by means of the formula[‡]

$$(1/2\pi)\int_{-\infty}^{\infty} e^{i\alpha x}\,\mathrm{d}x = \delta(\alpha). \tag{12.9}$$

We have

$$\begin{aligned}\int \psi_{\mathbf{p}'}\psi_{\mathbf{p}}^*\,\mathrm{d}V &= C^2\int e^{(i/\hbar)(\mathbf{p}'-\mathbf{p})\cdot\mathbf{r}}\,\mathrm{d}V\\ &= C^2(2\pi\hbar)^3\,\delta(\mathbf{p}'-\mathbf{p}).\end{aligned}$$

[†] The δ-function of a vector argument is defined as a product of δ-functions of the components of the vector.

[‡] The conventional meaning of this formula is that the integral on the left-hand side has the properties of the δ-function. When $\alpha = 0$, the integral diverges; when $\alpha \neq 0$, it is zero, being the integral of a periodic function with alternating sign. A further integration over α from $-L$ to $+L$, the range including the point $\alpha = 0$, gives

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-L}^{L}e^{i\alpha x}\,\mathrm{d}\alpha = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin Lx}{x}\,\mathrm{d}x = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin\xi}{\xi}\,\mathrm{d}\xi = 1.$$

Hence we see that we must have $C^2(2\pi\hbar)^3 = 1$. Thus the normalised function is

$$\psi_{\mathbf{p}} = (2\pi\hbar)^{-3/2}e^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}}. \tag{12.10}$$

The expansion of an arbitrary wave function $\psi(\mathbf{r})$ of a particle in terms of the eigenfunctions $\psi_{\mathbf{p}}$ of its momentum operator is simply the expansion as a Fourier integral:

$$\psi(\mathbf{r}) = \int a(\mathbf{p})\,\psi_{\mathbf{p}}(\mathbf{r})\,\mathrm{d}^3p = (2\pi\hbar)^{-3/2}\int a(\mathbf{p})e^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}}\,\mathrm{d}^3p \tag{12.11}$$

(where $\mathrm{d}^3p = \mathrm{d}p_x\,\mathrm{d}p_y\,\mathrm{d}p_z$). The expansion coefficients $a(\mathbf{p})$ are, according to formula (5.2),

$$a(\mathbf{p}) = \int \psi(\mathbf{r})\,\psi^*_{\mathbf{p}}(\mathbf{r})\,\mathrm{d}V = (2\pi\hbar)^{-3/2}\int \psi(\mathbf{r})e^{-(i/\hbar)\mathbf{p}\cdot\mathbf{r}}\,\mathrm{d}V. \tag{12.12}$$

The function $a(\mathbf{p})$ can be regarded (see §5) as the wave function of the particle in the *momentum representation*; $|a(\mathbf{p})|^2\,\mathrm{d}^3p$ is the probability that the momentum has a value in the interval d^3p. The formulae (12.11) and (12.12) give the relation between the wave functions in the two representations.

§13. Uncertainty relations

Let us derive the rules for commutation between momentum and coordinate operators. Since the result of successively differentiating with respect to one of the variables x, y, z and multiplying by another of them does not depend on the order of these operations, we have

$$\hat{p}_x y - y\hat{p}_x = 0, \quad \hat{p}_x z - z\hat{p}_x = 0, \tag{13.1}$$

and similarly for $\hat{p}_y$, $\hat{p}_z$.

To derive the commutation rule for $\hat{p}_x$ and x, we write

$$\begin{aligned}(\hat{p}_x x - x\hat{p}_x)\psi &= -i\hbar\,\partial(x\psi)/\partial x + i\hbar x\,\partial\psi/\partial x\\ &= -i\hbar\psi.\end{aligned}$$

We see that the result of the action of the operator $\hat{p}_x x - x\hat{p}_x$ reduces to multiplication by $-i\hbar$; the same is true, of course, of the commutation of $\hat{p}_y$ with y and $\hat{p}_z$ with z. Thus we have[†]

$$\hat{p}_x x - x\hat{p}_x = -i\hbar, \quad \hat{p}_y y - y\hat{p}_y = -i\hbar, \quad \hat{p}_z z - z\hat{p}_z = -i\hbar. \qquad (13.2)$$

The relations (13.1) and 13.2) show that the coordinate of a particle along one of the axes can have a definite value at the same time as the components of the momentum along the other two axes; the coordinate and momentum component along the same axis, however, cannot exist simultaneously. In particular, the particle cannot be at a definite point in space and at the same time have a definite momentum **p**.

Let us suppose that the particle is in some finite region of space, whose dimensions along the three axes are (of the order of magnitude of) $\Delta x, \Delta y, \Delta z$. Also, let the mean value of the momentum of the particle be $\mathbf{p}_0$. Mathematically, this means that the wave function has the form $\psi = u(\mathbf{r})e^{(i/\hbar)\mathbf{p}_0\cdot\mathbf{r}}$, where $u(\mathbf{r})$ is a function which differs considerably from zero only in the region of space concerned. We expand the function ψ in terms of the eigenfunctions of the momentum operator (i.e. as a Fourier integral). The coefficients $a(\mathbf{p})$ in this expansion are determined by the integrals (12.12) of functions of the form $u(\mathbf{r})e^{(i/\hbar)(\mathbf{p}_0-\mathbf{p})\cdot\mathbf{r}}$. If this integral is to differ considerably from zero, the periods of the oscillatory factor $e^{(i/\hbar)(\mathbf{p}_0-\mathbf{p})\cdot\mathbf{r}}$ must not be small in comparison with the dimensions $\Delta x, \Delta y, \Delta z$ of the region in which the function $u(\mathbf{r})$ is different from zero. This means that $a(\mathbf{p})$ will be considerably different from zero only for values of **p** such that $(p_{0x}-p_x)\Delta x/\hbar \lesssim 1$, etc. Since $|a(\mathbf{p})|^2$ determines the probability of the various values of the momentum, the ranges of values of p_x, p_y, p_z in which $a(\mathbf{p})$ differs from zero are just those in which the components of the momentum of the particle may be found, in the state considered. Denoting these ranges by $\Delta p_x, \Delta p_y, \Delta p_z$, we thus have

$$\Delta p_x \Delta x \sim \hbar, \quad \Delta p_y \Delta y \sim \hbar, \quad \Delta p_z \Delta z \sim \hbar. \qquad (13.3)$$

These relations, known as the *uncertainty relations*, were obtained by Heisenberg in 1927.

[†] These relations, discovered in matrix form by Heisenberg in 1925, formed the genesis of modern quantum mechanics.

We see that, the greater the accuracy with which the coordinate of the particle is known (i.e. the less Δx), the greater the uncertainty Δp_x in the component of the momentum along the same axis, and vice versa. In particular, if the particle is at some completely definite point in space ($\Delta x = \Delta y = \Delta z = 0$), then $\Delta p_x = \Delta p_y = \Delta p_z = \infty$. This means that all values of the momentum are equally probable. Conversely, if the particle has a completely definite momentum $\mathbf{p}$, then all positions of it in space are equally probable (this is seen directly from the wave function (12.7), whose squared modulus is quite independent of the coordinates).

§14. Angular momentum

In §12, to derive the law of conservation of momentum, we have made use of the homogeneity of space relative to a closed system of particles. Besides its homogeneity, space has also the property of isotropy: all directions in it are equivalent. Hence the Hamiltonian of a closed system cannot change when the system rotates as a whole through an arbitrary angle about an arbitrary axis. It is sufficient to require the fulfilment of this condition for an infinitely small rotation.

Let $\delta\boldsymbol{\varphi}$ be the vector of an infinitely small rotation, equal in magnitude to the angle $\delta\phi$ of the rotation and directed along the axis about which the rotation takes place. The changes $\delta\mathbf{r}_a$ (in the radius vectors $\mathbf{r}_a$ of the particles) in such a rotation are

$$\delta\mathbf{r}_a = \delta\boldsymbol{\varphi}\times\mathbf{r}_a$$

(see *Mechanics and Electrodynamics*, §9). An arbitrary function $\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots)$ is thereby transformed into the function

$$\psi(\mathbf{r}_1+\delta\mathbf{r}_1, \mathbf{r}_2+\delta\mathbf{r}_2, \ldots) = \psi(\mathbf{r}_1, \mathbf{r}_2, \ldots)+\sum_a \delta\mathbf{r}_a\cdot\nabla_a\psi$$

$$= \left(1+\delta\boldsymbol{\varphi}\cdot\sum_a \mathbf{r}_a\times\nabla_a\right)\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots).$$

The expression

$$1+\delta\boldsymbol{\varphi}\,.\sum_a \mathbf{r}_a\times\nabla_a \tag{14.1}$$

can be regarded as the "operator of an infinitely small rotation". The fact that an infinitely small rotation does not alter the Hamiltonian of the system is expressed by the commutability of the rotation operator with the operator $\hat{H}$. Since $\delta\boldsymbol{\varphi}$ is a constant vector, this condition reduces to the relation

$$\Big(\sum_a \mathbf{r}_a\times\nabla_a\Big)\hat{H}-\hat{H}\Big(\sum_a \mathbf{r}_a\times\nabla_a\Big) = 0, \tag{14.2}$$

which expresses a certain law of conservation.

The quantity whose conservation for a closed system follows from the property of isotropy of space is the *angular momentum* of the system (see *Mechanics and Electrodynamics*, §9). Thus the operator $\Sigma\mathbf{r}_a\times\nabla_a$ must correspond, apart from a constant factor, to the total angular momentum of the system, and each of the terms $\mathbf{r}_a\times\nabla_a$ of this sum corresponds to the angular momentum of an individual particle.

The coefficient of proportionality must be put equal to $-i\hbar$; then the expression for the angular momentum operator of a particle is $-i\hbar\mathbf{r}\times\nabla = \mathbf{r}\times\hat{\mathbf{p}}$ and corresponds exactly to the familiar classical expression $\mathbf{r}\times\mathbf{p}$. Henceforward we shall always use the angular momentum measured in units of $\hbar$. The angular momentum operator of a particle, so defined, will be denoted by $\hat{\mathbf{l}}$, and that of the whole system by $\hat{\mathbf{L}}$. Thus we have for the angular momentum operator of a particle

$$\hbar\hat{\mathbf{l}} = \mathbf{r}\times\hat{\mathbf{p}} = -i\hbar\mathbf{r}\times\nabla,$$

or, in components,

$$\hbar\hat{l}_x = y\hat{p}_z-z\hat{p}_y,\quad \hbar\hat{l}_y = z\hat{p}_x-x\hat{p}_z,\quad \hbar\hat{l}_z = x\hat{p}_y-y\hat{p}_x. \tag{14.3}$$

For a system which is in an external field, the angular momentum is in general not conserved. However, it may still be conserved if the field has a certain symmetry. Thus, if the system is in a centrally symmetric field, all directions in space at the centre are equivalent, and hence the angular momentum about this centre will be conserved. Similarly, in an axially symmetric field, the component of angular momentum along the axis of symmetry is conserved. All these conservation laws holding in classical mechanics are valid in quantum mechanics also.

Let us derive the rules for commutation of the angular momentum operators with those of coordinates and linear momenta. For instance,

$$\begin{aligned}\hat{l}_x y - y\hat{l}_x &= (1/\hbar)(y\hat{p}_z - z\hat{p}_y)y - y(y\hat{p}_z - z\hat{p}_y)(1/\hbar)\\ &= -(1/\hbar)\,z(\hat{p}_y y - y\hat{p}_y) = iz.\end{aligned}$$

Similarly, we find

$$\hat{l}_x x - x\hat{l}_x = 0, \quad \hat{l}_x y - y\hat{l}_x = iz, \quad \hat{l}_x z - z\hat{l}_x = -iy, \tag{14.4}$$

and two other sets of three relations obtained from these by cyclic interchange of the coordinates and suffixes x, y, z.

It is easily seen that similar commutation rules hold for the angular momentum and linear momentum operators:

$$\hat{l}_x\hat{p}_x - \hat{p}_x\hat{l}_x = 0, \quad \hat{l}_x\hat{p}_y - \hat{p}_y\hat{l}_x = i\hat{p}_z, \quad \hat{l}_x\hat{p}_z - \hat{p}_z\hat{l}_x = -i\hat{p}_y. \tag{14.5}$$

By means of these formulae, it is easy to find the rules for commutation of the operators $\hat{l}_x$, $\hat{l}_y$, $\hat{l}_z$ with one another. We have

$$\begin{aligned}\hbar(\hat{l}_x\hat{l}_y - \hat{l}_y\hat{l}_x) &= \hat{l}_x(z\hat{p}_x - x\hat{p}_z) - (z\hat{p}_x - x\hat{p}_z)\hat{l}_x\\ &= (\hat{l}_x z - z\hat{l}_x)\hat{p}_x - x(\hat{l}_x\hat{p}_z - \hat{p}_z\hat{l}_x)\\ &= -iy\hat{p}_x + ix\hat{p}_y = i\hbar\hat{l}_z.\end{aligned}$$

Thus

$$\hat{l}_y\hat{l}_z - \hat{l}_z\hat{l}_y = i\hat{l}_x, \quad \hat{l}_z\hat{l}_x - \hat{l}_x\hat{l}_z = i\hat{l}_y, \quad \hat{l}_x\hat{l}_y - \hat{l}_y\hat{l}_x = i\hat{l}_z. \tag{14.6}$$

Exactly similar relations hold for the operators $\hat{L}_x$, $\hat{L}_y$, $\hat{L}_z$ of the total angular momentum of the system. For, since the angular momentum operators of different individual particles commute, we have, for instance,

$$\sum_a \hat{l}_{ay} \sum_a \hat{l}_{az} - \sum_a \hat{l}_{az} \sum_a \hat{l}_{ay} = \sum_a (\hat{l}_{ay}\hat{l}_{az} - \hat{l}_{az}\hat{l}_{ay}) = i \sum_a \hat{l}_{ax}.$$

Thus

$$\hat{L}_y\hat{L}_z - \hat{L}_z\hat{L}_y = i\hat{L}_x, \quad \hat{L}_z\hat{L}_x - \hat{L}_x\hat{L}_z = i\hat{L}_y, \quad \hat{L}_x\hat{L}_y - \hat{L}_y\hat{L}_x = i\hat{L}_z. \tag{14.7}$$

The relations (14.7) show that the three components of the angular momentum cannot simultaneously have definite values (except in the case where all three components simultaneously vanish: see below). In this respect the angular momentum is fundamentally different from the linear momentum, whose three components can be simultaneously measured.

From the operators $\hat{L}_x, \hat{L}_y, \hat{L}_z$ we can form the operator of the square of the modulus of the angular momentum vector:

$$\hat{\mathbf{L}}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2. \tag{14.8}$$

This operator commutes with each of the operators $\hat{L}_x, \hat{L}_y, \hat{L}_z$:

$$\hat{\mathbf{L}}^2\hat{L}_x - \hat{L}_x\hat{\mathbf{L}}^2 = 0, \quad \hat{\mathbf{L}}^2\hat{L}_y - \hat{L}_y\hat{\mathbf{L}}^2 = 0, \quad \hat{\mathbf{L}}^2\hat{L}_z - \hat{L}_z\hat{\mathbf{L}}^2 = 0. \tag{14.9}$$

Using (14.7), we have, for example,

$$\begin{aligned}
\hat{L}_x^2\hat{L}_z - \hat{L}_z\hat{L}_x^2 &= \hat{L}_x(\hat{L}_x\hat{L}_z - \hat{L}_z\hat{L}_x) + (\hat{L}_x\hat{L}_z - \hat{L}_z\hat{L}_x)\hat{L}_x \\
&= -i(\hat{L}_x\hat{L}_y + \hat{L}_y\hat{L}_x), \\
\hat{L}_y^2\hat{L}_z - \hat{L}_z\hat{L}_y^2 &= i(\hat{L}_x\hat{L}_y + \hat{L}_y\hat{L}_x), \\
\hat{L}_z^2\hat{L}_z - \hat{L}_z\hat{L}_z^2 &= 0.
\end{aligned}$$

Adding these equations, we have the third relation (14.9). Physically, the relations (14.9) mean that the square of the angular momentum, i.e. its modulus, can have a definite value at the same time as one of its components.

Instead of the operators $\hat{L}_x, \hat{L}_y$ it is often more convenient to use the complex combinations

$$\hat{L}_+ = \hat{L}_x + i\hat{L}_y, \quad \hat{L}_- = \hat{L}_x - i\hat{L}_y. \tag{14.10}$$

It is easily verified by direct calculation using (14.7) that the following commutation rules hold:

$$\begin{gathered} \hat{L}_+\hat{L}_- - \hat{L}_-\hat{L}_+ = 2\hat{L}_z, \quad \hat{L}_z\hat{L}_+ - \hat{L}_+\hat{L}_z = \hat{L}_+, \\ \hat{L}_z\hat{L}_- - \hat{L}_-\hat{L}_z = -\hat{L}_-, \end{gathered} \tag{14.11}$$

and it is also not difficult to see that

$$\hat{\mathbf{L}}^2 = \hat{L}_-\hat{L}_+ + \hat{L}_z^2 + \hat{L}_z. \tag{14.12}$$

Finally, we shall give some frequently used expressions for the angular momentum operator of a single particle in spherical polar coordinates. Defining the latter by means of the usual relations

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\theta,$$

we have after a simple calculation

$$\hat{l}_z = -i\frac{\partial}{\partial\phi}, \tag{14.13}$$

$$\hat{l}_\pm = e^{\pm i\phi}\left(\pm\frac{\partial}{\partial\theta} + i\cot\theta\,\frac{\partial}{\partial\phi}\right). \tag{14.14}$$

Substitution in (14.12) gives the squared angular momentum operator of the particle:

$$\hat{\mathbf{l}}^2 = -\left[\frac{1}{\sin^2\theta}\,\frac{\partial^2}{\partial\phi^2} + \frac{1}{\sin\theta}\,\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial}{\partial\theta}\right)\right]. \tag{14.15}$$

It should be noticed that this is, apart from a factor, the angular part of the Laplacian operator.

§15. Eigenvalues of the angular momentum

In order to determine the eigenvalues of the component, in some direction, of the angular momentum of a particle, it is convenient to use the expression for its operator in spherical polar coordinates, taking the direction in question as the polar axis. According to formula (14.13), the equation $\hat{l}_z\psi = l_z\psi$ can be written in the form

$$-i\,\partial\psi/\partial\phi = l_z\psi. \tag{15.1}$$

Its solution is

$$\psi = f(r,\theta)\,e^{il_z\phi},$$

where $f(r,\theta)$ is an arbitrary function of r and θ. If the function ψ is to be single-valued, it must be periodic in ϕ, with period 2π. Hence we find[†]

$$l_z = m, \quad \text{where} \quad m = 0, \pm 1, \pm 2, \ldots. \tag{15.2}$$

Thus the eigenvalues l_z are the positive and negative integers, including zero. The factor depending on ϕ, which characterises the eigenfunctions of the operator $\hat{l}_z$, is denoted by

$$\Phi_m(\phi) = (2\pi)^{-1/2}e^{im\phi}. \tag{15.3}$$

These functions are normalised so that

$$\int_0^{2\pi} \Phi_m^*(\phi)\,\Phi_{m'}(\phi)\,d\phi = \delta_{mm'}. \tag{15.4}$$

The eigenvalues of the z-component of the total angular momentum of the system are evidently also equal to the positive and negative integers:

$$L_z = M, \quad \text{where} \quad M = 0, \pm 1, \pm 2, \ldots \tag{15.5}$$

(this follows at once from the fact that the operator $\hat{L}_z$ is equal to the sum of the commuting operators $\hat{l}_z$ for the individual particles).

[†] The customary notation for the eigenvalues of the angular momentum component is m, which also denotes the mass of a particle, but this should not lead to any confusion.

Since the direction of the z-axis is in no way distinctive, it is clear that the same result is obtained for L_x, L_y and in general for the component of the angular momentum in any direction: they can all take integral values only. At first sight this result may appear paradoxical, particularly if we apply it to two directions infinitely close to each other. In fact, however, it must be remembered that the only common eigenfunction of the operators $\hat{L}_x$, $\hat{L}_y$, $\hat{L}_z$ corresponds to the simultaneous values

$$L_x = L_y = L_z = 0;$$

in this case the angular momentum vector is zero, and consequently so is its projection upon any direction. If even one of the eigenvalues L_x, L_y, L_z is not zero, the operators $\hat{L}_x$, $\hat{L}_y$, $\hat{L}_z$ have no common eigenfunctions. In other words, there is no state in which two or three of the angular momentum components in different directions simultaneously have definite values different from zero, so that we can say only that one of them is integral.

The stationary states of a system which differ only in the value of M have the same energy; this follows from general considerations, based on the fact that the direction of the z-axis is in no way distinctive. Thus the energy levels of a system whose angular momentum is conserved (and is not zero) are always degenerate.[†]

Let us now look for the eigenvalues of the square of the angular momentum. We shall show how these values may be found, starting from the commutation rules (14.7) only. We denote by ψ_M the wave functions of the stationary states belonging to one degenerate energy level and having the same value of $\mathbf{L}^2$.

First of all we note that, since the two directions of the z-axis are physically equivalent, for every possible positive value $M = +|M|$ there is a corresponding negative value $M = -|M|$. Let L (a positive integer) denote the greatest possible value of $|M|$.

[†] This is a particular case of the general theorem, mentioned in §10, which states that the levels are degenerate when two or more conserved quantities exist whose operators do not commute. Here the components of the angular momentum are such quantities.

On applying the operator $\hat{L}_z\hat{L}_\pm$ to the eigenfunction ψ_M of the operator $\hat{L}_z$ and using the commutation rule (14.11), we find

$$\hat{L}_z\hat{L}_\pm\psi_M = \hat{L}_\pm\hat{L}_z\psi_M \pm \hat{L}_\pm\psi_M = (M\pm 1)\,\hat{L}_\pm\psi_M.$$

Hence we see that the function $\hat{L}_\pm\psi_M$ is (apart from a normalisation constant) the eigenfunction corresponding to the value $M\pm 1$ of the quantity L_z; we can write

$$\psi_{M+1} = \text{constant}\times\hat{L}_+\psi_M, \quad \psi_{M-1} = \text{constant}\times\hat{L}_-\psi_M. \tag{15.6}$$

If we put $M = L$ in the first of these equations, we must have identically

$$\hat{L}_+\psi_L = 0, \tag{15.7}$$

since there is by definition no state with $M > L$. Applying the operator $\hat{L}_-$ to this equation and using the relation (14.12), we obtain

$$\hat{L}_-\hat{L}_+\psi_L = (\hat{\mathbf{L}}^2 - \hat{L}_z^2 - \hat{L}_z)\,\psi_L = 0.$$

Since, however, the ψ_M are common eigenfunctions of the operators $\hat{\mathbf{L}}^2$ and $\hat{L}_z$, we have

$$\hat{\mathbf{L}}^2\psi_L = \mathbf{L}^2\psi_L, \quad \hat{L}_z^2\psi_L = L^2\psi_L, \quad \hat{L}_z\psi_L = L\psi_L,$$

so that the equation found above gives

$$\mathbf{L}^2 = L(L+1). \tag{15.8}$$

This formula determines the required eigenvalues of the square of the angular momentum; the number L takes all positive integral values, including zero. For a given value of L, the component $L_z = M$ of the angular momentum can take the values

$$M = L, L-1, \ldots, -L, \tag{15.9}$$

i.e. $2L+1$ different values in all. The energy level corresponding to the angular momentum L thus has $(2L+1)$-fold degeneracy. This is usually called *degeneracy with respect to the direction of the angular momentum.* A state with angular momentum $L = 0$ (when all three components are also zero) is not degenerate; we notice that the wave function of such a state is spherically symmetric. This is clear from the fact that the change in the wave function in any infinitely small rotation, given by $\hat{\mathbf{L}}\psi$, is in this case zero.

We shall often, for the sake of brevity and in accordance with custom, speak of the "angular momentum" L of a system, understanding by this an angular momentum whose square is $L(L+1)$; the angular momentum of a single particle will be denoted by the lowercase letter l. The z-component of the angular momentum is usually called just the "angular-momentum component".

Let us calculate the matrix elements of the quantities L_x and L_y for transitions between states having the same energy and angular momentum L, but different values of the angular-momentum component M.

It is seen from formulae (15.6) that, in the matrices of the operators $\hat{L}_+$ and $\hat{L}_-$, only those elements are different from zero which correspond to transitions $M \to M+1$ and $M \to M-1$ respectively. Taking this into account, we find the diagonal matrix elements (for transitions $L, M-1 \to L, M-1$) on both sides of the equation (14.12), obtaining

$$L(L+1) = (L_-)_{M-1,M}(L_+)_{M,M-1}+M^2-M.$$

Noticing that, since the operators $\hat{L}_x$ and $\hat{L}_y$ are Hermitian,

$$(L_-)_{M-1,M} = (L_+)^*_{M,M-1},$$

we can rewrite this equation in the form

$$|(L_+)_{M,M-1}|^2 = L(L+1)-M(M-1) = (L-M+1)(L+M),$$

whence (with the notation (11.3))

$$\langle M|L_+|M-1\rangle = \langle M-1|L_-|M\rangle = \sqrt{[(L+M)(L-M+1)]}. \tag{15.10}$$

Hence we have for the non-zero matrix elements of the quantities L_x and L_y themselves

$$\begin{aligned}\langle M|L_x|M-1\rangle &= \langle M-1|L_x|M\rangle \\ &= \tfrac{1}{2}\sqrt{[(L+M)(L-M+1)]}, \\ \langle M|L_y|M-1\rangle &= -\langle M-1|L_y|M\rangle \\ &= -\tfrac{1}{2}i\sqrt{[(L+M)(L-M+1)]}.\end{aligned} \tag{15.11}$$

The diagonal elements in the matrices of the quantities L_x and L_y are zero. Since the diagonal matrix element gives the mean value of the quantity in the corresponding state, it follows that the mean values $\overline{L_x}$ and $\overline{L_y}$ are zero in states having definite values of L_z. Thus, if the angular-momentum component in a given direction in space has a definite value, the vector $\overline{\mathbf{L}}$ itself is in that direction.

§16. Eigenfunctions of the angular momentum

The wave function of a particle is not completely determined when the values of l and m are prescribed. This is seen from the fact that the expressions for the operators of these quantities in spherical polar coordinates contain only the angles θ and ϕ, so that their eigenfunctions can contain an arbitrary factor depending on r. We shall here consider only the angular part of the wave function which characterises the eigenfunctions of the angular momentum, and denote this by $Y_{lm}(\theta, \phi)$, with the normalisation condition

$$\int |Y_{lm}|^2 \, \mathrm{d}o = 1,$$

where $\mathrm{d}o = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$ is an element of solid angle.

The functions Y_{lm} with different l or m are automatically orthogonal, as being the eigenfunctions of angular momentum operators corresponding to different eigenvalues, and together with the normalisation condition this gives

$$\int_0^{2\pi}\int_0^{\pi} Y^*_{l'm'} Y_{lm} \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi = \delta_{ll'}\delta_{mm'}. \qquad (16.1)$$

The most direct method of calculating the required functions is by directly solving the problem of finding the eigenfunctions of the operator $\hat{\mathbf{l}}^2$ written in spherical polar coordinates. The equation

$$\hat{\mathbf{l}}^2 \psi = l(l+1)\psi$$

becomes

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial\psi}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2} + l(l+1)\psi = 0. \qquad (16.2)$$

This equation admits of separation of the variables, and its solution can be sought in the form

$$Y_{lm} = \Phi_m(\phi)\, \Theta_{lm}(\theta), \tag{16.3}$$

where the Φ_m are the functions (15.3). Substituting (16.3) in (16.2), we obtain for the function Θ_{lm} the equation

$$\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\,\frac{\mathrm{d}\Theta_{lm}}{\mathrm{d}\theta}\right) - \frac{m^2}{\sin^2\theta}\Theta_{lm} + l(l+1)\,\Theta_{lm} = 0. \tag{16.4}$$

This equation is well known in the theory of spherical harmonics. It has solutions satisfying the conditions of finiteness and single-valuedness for positive integral values of $l \geqslant |m|$, in agreement with the eigenvalues of the angular momentum obtained above by the matrix method. The corresponding solutions are what are called *associated Legendre polynomials* $P_l^m(\cos\theta)$.

Thus the angular wave functions are

$$Y_{lm}(\theta, \phi) = \text{constant} \times P_l^m(\cos\theta)\, e^{im\phi}, \tag{16.5}$$

and are mathematically just spherical harmonic functions normalised in a particular way. We shall not give here the general expression for the normalisation constant, but only list the explicit formulae for the first few functions ($l = 0, 1, 2$):

$$\left.\begin{aligned} Y_{00} &= 1/\sqrt{(4\pi)}, \\ Y_{10} &= \sqrt{(3/4\pi)}\cos\theta, \quad Y_{1,\pm1} = \mp\sqrt{(3/8\pi)}\sin\theta\, e^{\pm i\phi}, \\ Y_{20} &= \sqrt{(5/16\pi)}\,(3\cos^2\theta - 1), \\ Y_{2,\pm1} &= \mp\sqrt{(15/8\pi)}\cos\theta\sin\theta\, e^{\pm i\phi}, \\ Y_{2,\pm2} &= \sqrt{(15/32\pi)}\sin^2\theta\, e^{\pm 2i\phi}. \end{aligned}\right\} \tag{16.6}$$

For $m = 0$, the associated Legendre polynomials are called simply *Legendre polynomials* $P_l(\cos\theta)$. The corresponding normalised spherical harmonics are

$$Y_{l0} = \sqrt{\frac{2l+1}{4\pi}}\, P_l(\cos\theta). \tag{16.7}$$

If $l = 0$ (so that $m = 0$ also), the function (16.7) reduces to a constant. In other words, the wave functions of the states of a particle with $l = 0$ depend only on r, i.e. they have complete spherical symmetry, in accordance with the general result given in §15. If one of the spherical harmonics in (16.1) is Y_{00}, then the other satisfies

$$\int Y_{lm}\, do = 0 \quad (l \neq 0). \tag{16.8}$$

§17. Addition of angular momenta

Let us consider a system composed of two parts whose interaction is weak. If the interaction is entirely neglected, then for each part the law of conservation of angular momentum holds. The angular momentum $\mathbf{L}$ of the whole system can be regarded as the sum of the angular momenta $\mathbf{L}_1$ and $\mathbf{L}_2$ of its parts. In the next approximation, when the weak interaction is taken into account, $\mathbf{L}_1$ and $\mathbf{L}_2$ are not exactly conserved, but the numbers L_1 and L_2 which determine their squares remain "good" quantum numbers suitable for an approximate description of the state of the system.

For such systems the question arises regarding the "law of addition" of angular momenta: what are the possible values of L for given values of L_1 and L_2? The law of addition for the components of angular momentum is evident: since $\hat{L}_z = \hat{L}_{1z} + \hat{L}_{2z}$, it follows that

$$M = M_1 + M_2. \tag{17.1}$$

There is no such simple relation for the operators of the squared angular momenta, however, and to derive their "law of addition" we reason as follows.

If we take the quantities $\mathbf{L}_1^2$, $\mathbf{L}_2^2$, L_{1z}, L_{2z}, as a complete set of physical quantities,[†] every state will be determined by the values of the numbers L_1, L_2, M_1, M_2. For given L_1 and L_2, the numbers M_1 and

---

[†] Together with such other quantities as form a complete set when combined with these four. These other quantities play no part in the subsequent discussion, and for brevity we shall ignore them entirely, and conventionally call the above four quantities a complete set.

M_2 take $(2L_1+1)$ and $(2L_2+1)$ different values respectively, so that there are altogether $(2L_1+1)(2L_2+1)$ different states with the same L_1 and L_2. We denote the wave functions of the states for this representation by $\phi_{L_1L_2M_1M_2}$.

Instead of the above four quantities, we can take the four quantities $\mathbf{L}_1^2$, $\mathbf{L}_2^2$, $\mathbf{L}^2$, L_z as a complete set. Then every state is characterised by the values of the numbers L_1, L_2, L, M (we denote the corresponding wave functions by $\psi_{L_1L_2LM}$). For given L_1 and L_2, there must of course be $(2L_1+1)(2L_2+1)$ different states as before, i.e. for given L_1 and L_2 the pair of numbers L and M must take $(2L_1+1)(2L_2+1)$ pairs of values. These values can be determined as follows.

By adding the various possible values of M_1 and M_2, we get the corresponding values of M, as shown below:

M_1	M_2	M
L_1	L_2	L_1+L_2
L_1	L_2-1	L_1+L_2-1
L_1-1	L_2	
L_1	L_2-2	L_1+L_2-2
L_1-1	L_2-1	
L_1-2	L_2	
...	...	...

We see that the greatest possible value of M is $M = L_1+L_2$, corresponding to one state ϕ (one pair of values of M_1 and M_2). The greatest possible value of M in the states ψ, and hence the greatest possible value of L also, is therefore L_1+L_2. Next, there are two states ϕ with $M = L_1+L_2-1$. Consequently, there must also be two states ψ with this value of M; one of them is the state with $L = L_1+L_2$ (and $M = L-1$), and the other is that with $L = L_1+L_2-1$ (and $M = L$). For the value $M = L_1+L_2-2$ there are three different states ϕ. This means that, besides the values $L = L_1+L_2$, $L = L_1+L_2-1$, the value $L = L_1+L_2-2$ can occur.

The argument can be continued in this way so long as a decrease of M by 1 increases by 1 the number of states with a given M. It is easily seen that this is so until M reaches the value $|L_1-L_2|$. When M decreases further, the number of states no longer increases, remaining equal to $2L_2+1$ (if $L_2 \leqslant L_1$). Thus $|L_1-L_2|$ is the least possible value of L, and we arrive at the result that, for given L_1 and L_2, the number L can take the values

$$L = L_1+L_2, L_1+L_2-1, \ldots, |L_1-L_2|, \tag{17.2}$$

that is $2L_2+1$ different values altogether (supposing that $L_2 \leqslant L_1$). It is easy to verify that we do in fact obtain $(2L_1+1)(2L_2+1)$ different values of the pair of numbers L, M. Here it is important to note that, if we ignore the $2L+1$ values of M for a given L, then only one state will correspond to each of the possible values (17.2) of L.

This result can be illustrated by means of what is called the *vector model*. If we take two vectors $\mathbf{L}_1$ and $\mathbf{L}_2$ of lengths L_1 and L_2, then the values of L are represented by the integral lengths of the vector $\mathbf{L}$ which are obtained by vector addition of $\mathbf{L}_1$ and $\mathbf{L}_2$; the greatest value of L is L_1+L_2, which is obtained when $\mathbf{L}_1$ and $\mathbf{L}_2$ are parallel, and the least value is $|L_1-L_2|$, when $\mathbf{L}_1$ and $\mathbf{L}_2$ are antiparallel.

In states with definite values of the angular momenta L_1, L_2 and of the total angular momentum L, the scalar products $\mathbf{L}_1\cdot\mathbf{L}_2$, $\mathbf{L}\cdot\mathbf{L}_1$ and $\mathbf{L}\cdot\mathbf{L}_2$ also have definite values. These values are easily found. To calculate $\mathbf{L}_1\cdot\mathbf{L}_2$, we write $\hat{\mathbf{L}} = \hat{\mathbf{L}}_1+\hat{\mathbf{L}}_2$ or, squaring and transposing,

$$2\hat{\mathbf{L}}_1\cdot\hat{\mathbf{L}}_2 = \hat{\mathbf{L}}^2-\hat{\mathbf{L}}_1^2-\hat{\mathbf{L}}_2^2.$$

Replacing the operators on the right-hand side of this equation by their eigenvalues, we obtain the eigenvalue of the operator on the left-hand side:

$$\mathbf{L}_1\cdot\mathbf{L}_2 = \tfrac{1}{2}\{L(L+1)-L_1(L_1+1)-L_2(L_2+1)\}. \tag{17.3}$$

Similarly we find

$$\mathbf{L}\cdot\mathbf{L}_1 = \tfrac{1}{2}\{L(L+1)+L_1(L_1+1)-L_2(L_2+1)\}. \tag{17.4}$$

If $\psi^{(1)}_{L_1M_1}$ and $\psi^{(2)}_{L_2M_2}$ are the wave functions of the two parts of the system, then the wave function of the whole system (the interaction of the parts being again neglected) is the product

$$\phi_{L_1L_2M_1M_2} = \psi^{(1)}_{L_1M_1}\psi^{(2)}_{L_2M_2}. \tag{17.5}$$

These states have definite values of M_1 and M_2 (as well as of L_1 and L_2). The states having definite values of L and M are superpositions of the states (17.5) with various values of M_1 and M_2 such that $M_1+M_2 = M$. Their wave functions are the linear combinations

$$\psi_{L_1L_2LM} = \sum_{M_1,\,M_2} C^{L_1L_2LM}_{L_1L_2M_1M_2}\phi_{L_1L_2M_1M_2} \tag{17.6}$$

with coefficients C which depend on the values of all the quantum numbers shown. These coefficients are called *vector addition coefficients* or *Clebsch–Gordan coefficients.*

§18. **Angular momentum selection rules**

We have seen that in both classical and quantum mechanics the law of conservation of angular momentum is a consequence of the isotropy of space with respect to a closed system. This already demonstrates the relation between the angular momentum and the symmetry properties under rotation. But in quantum mechanics the relation in question is a particularly far-reaching one, and essentially constitutes the basic content of the concept of angular momentum, especially as the classical definition of the angular momentum of a particle as the product $\mathbf{r}\times\mathbf{p}$ has no direct significance in quantum mechanics, the vectors of momentum and angular momentum not being simultaneously measurable.

It has been shown in §16 that, if the values of l and m are specified, the angular dependence of the wave function of the particle is determined, and therefore so are all its symmetry properties under rotation. The most general formulation of these properties involves specifying the transformation of the wave functions when the coordinate system is rotated.

The wave function ψ_{LM} of a system of particles (with specified values of L and M) remains constant[†] only under a rotation of the coordinate system about the z-axis. Any rotation that alters the direction of the z-axis has the result that the angular momentum component along the new z-axis does not have a definite value. This means that, in the new coordinates, the wave function in general becomes a superposition (a linear combination) of $2L+1$ functions corresponding to the possible values of M for the given L. We can say that the $2L+1$ functions ψ_{LM} are transformed into linear combinations of one another when the coordinate system is rotated.[‡] The law governing this transformation (i.e. the coefficients in the linear combination, as functions of the angles of rotation of the coordinate axes) is entirely determined by specifying the value of L. Thus the angular momentum L acquires the significance of a quantum number which classifies the states of the system according to their transformation properties under rotation of the coordinate system. This aspect of the concept of angular momentum in quantum mechanics is particularly important because it is not directly related to the explicit angular dependence of the wave functions; the law of mutual transformation of these functions can be stated without reference to that dependence.

We shall show how this approach can be used to find the *selection rules* (with respect to angular momentum) for the matrix elements of various quantities, i.e. the rules that determine the transitions for which the matrix elements can be different from zero.

To do so, we first note that the concept of angular momentum as a classificatory feature can be applied, in a purely conventional mathematical manner, not only to the wave functions but also to other physical quantities. For example, the angular momentum $L = 0$ "corresponds" to any scalar quantity (i.e. a quantity which is unchanged by a transformation of the coordinates), in the sense that when $L = 0$, $2L+1 = 1$, i.e. there is only one quantity that is "transformed

---

[†] Apart from an unimportant phase factor.

[‡] In mathematical terms, these functions are said to form *irreducible representations of the rotation group*. The number of functions which are transformed into linear combinations of one another is called the *dimension* of the representation; it is assumed that this number cannot be made smaller by taking any other linear combinations of these functions.

into a linear combination of itself".[†] Similarly, the angular momentum $L = 1$ ($2L+1 = 3$) can be assigned to a vector quantity, since the three independent components of the vector are transformed into linear combinations of one another when the coordinate system is rotated. If the components of the vector are expressed in terms of the spherical polar angles θ and ϕ which define its direction, we have

$$\left.\begin{aligned} A_+ &\equiv A_x+iA_y = A \sin\theta\, e^{i\phi} && (M = 1),\\ A_- &\equiv A_x-iA_y = A \sin\theta\, e^{-i\phi} && (M = -1),\\ A_z &= A\cos\theta && (M = 0). \end{aligned}\right\} \qquad (18.1)$$

Comparison with the functions (16.6) shows that the component A_z corresponds to the angular momentum component $M = 0$, and the complex combinations A_+ and A_- correspond to $M = 1$ and -1 respectively.

For simplicity and greater clarity, we shall consider quantities characterising the states of a single particle (free, or in a centrally symmetric external field). Let f be any scalar physical quantity. Let us take its matrix elements with respect to states having definite values of l and m:

$$\langle n'l'm' \,|f|\, nlm\rangle = \int \psi^*_{l'm'}\hat{f}\psi_{lm}\, \mathrm{d}V, \qquad (18.2)$$

where n and n' are the remaining quantities (other than l and m) which define the states of the particle.

The three factors in the integrand ($\psi^*_{l'm'}$, f and ψ_{lm}) can be put in correspondence with pairs of values of the angular momentum and its component: (l', $-m'$), (0, 0), and (l, m); taking the complex conjugate changes the sign of the exponent $im\phi$ in (16.5), thus effectively changing the sign of the angular momentum component. These angular momenta can be added in each of the possible ways to form a total angular momentum and its component (denoted by Λ and μ). This shows the transformation properties of the functions of which a linear combina-

---

[†] To avoid confusion, it should be emphasised that from this standpoint the wave functions ψ_{LM} (with $L \neq 1$) are not "scalars"; all the $2L+1$ functions ψ_{LM} with various values of M are to be regarded in this respect as constituting a single quantity.

tion can in principle give an expansion of the integrand in (18.2):

$$\psi^*_{l'm'}\hat{f}\psi_{lm} = \sum_{\Lambda} a_{\Lambda\mu}\psi_{\Lambda\mu} \qquad (\mu = m-m'), \tag{18.3}$$

where the $a_{\Lambda\mu}$ are constants and the $\psi_{\Lambda\mu}$ are functions whose transformation properties are the same as those of the angular momentum eigenfunctions. To solve the problem of the selection rules, however, there is no need to carry out this expansion. We need only note that all the terms of the sum except that with $\Lambda = \mu = 0$ give zero on integration over angles (from the condition (16.8)). The matrix element (18.2) can therefore be different from zero only if the values $\Lambda = \mu = 0$ are actually present in the expansion (18.3), and the value $\Lambda = 0$ can be obtained on adding two angular momenta l and l' only if $l = l'$.

Thus we conclude that the matrix elements of a scalar can differ from zero only for transitions in which the angular momentum and its component are unchanged:

$$l' = l, \quad m' = m. \tag{18.4}$$

Moreover, since specifying m defines only the orientation of the system with respect to the coordinate axes, and the value of the scalar f is independent of this orientation, we can say that the matrix elements $\langle n'lm|f|nlm\rangle$ are independent of m.

The selection rules for the matrix elements $\langle n'l'm'|\mathbf{A}|nlm\rangle$ of a vector **A** can be found in a similar manner. The vector is assigned an "angular momentum" of 1. Adding this to the angular momentum l, we get the values $l+1, l, l-1$ (if $l \neq 0$; when $l = 0$, the addition gives only the single value 1). The subsequent addition of the angular momentum l' must give a total "angular momentum" $\Lambda = 0$ if the integral is to differ from zero. Then l' must be equal to one of the results of the previous addition, i.e. the permissible values are

$$l' = l, l\pm 1, \tag{18.5}$$

and the matrix elements for transitions between states with $l' = l = 0$ are also forbidden.

The selection rules for the angular momentum components *m* are different for the different components of the vector. Using (18.1), we can easily derive the following rules:

$$\left.\begin{array}{lll} \text{for} & A_+ = A_x + iA_y, & M' = M+1, \\ \text{for} & A_- = A_x - iA_y, & M' = M-1, \\ \text{for} & A_z, & M' = M. \end{array}\right\} \qquad (18.6)$$

The matrix elements of the vector quantity depend on the values of M. It can be shown (although we shall not pause to do so here) that this dependence also is universal, being an unequivocal consequence of the transformation properties of the angular momentum eigenfunctions.

One further case is that of a symmetrical tensor A_{ik} of rank two. Such a tensor has six different components, but these do not form a single group as regards their transformation properties. The reason is that the trace of the tensor (i.e. the sum $A_{ii} = A_{xx} + A_{yy} + A_{zz}$) is a scalar, and must be excluded from the set of quantities undergoing transformation; in other words, we must consider a tensor having zero trace. Such a tensor is said to be *irreducible*; it has five independent components, and can be assigned the angular momentum $L = 2$ $(2L+1 = 5)$.[†]

It must be emphasised that, although we have here referred only to matrix elements for a single particle, all the results are in fact consequences of the general transformation properties of the wave functions, and are therefore equally valid for any system of particles whose angular momentum is conserved.

§19. **Parity of a state**

Besides the parallel displacements and rotations of the coordinate system, the invariance under which represents the homogeneity and isotropy of space respectively, there is another transformation which

---

[†] An example of a physical quantity of this kind is the electric quadrupole moment of a system.

leaves unaltered the Hamiltonian of a closed system. This is what is called the *inversion transformation*, which consists in simultaneously changing the sign of all the coordinates, i.e. a reversal of the direction of each coordinate axis; a right-handed coordinate system then becomes left-handed, and vice versa. The invariance of the Hamiltonian under this transformation expresses the symmetry of space under mirror reflections.† In classical mechanics, the invariance of Hamilton's function with respect to inversion does not lead to a conservation law, but the situation is different in quantum mechanics.

Let us denote by $\hat{P}$ (for "parity") a symbolic inversion operator whose effect on a wave function $\psi(\mathbf{r})$ is to change the sign of the coordinates:

$$\hat{P}\psi(\mathbf{r}) = \psi(-\mathbf{r}). \qquad (19.1)$$

It is easy to find the eigenvalues P of this operator, which are determined by the equation

$$\hat{P}\psi(\mathbf{r}) = P\psi(\mathbf{r}). \qquad (19.2)$$

To do this, we notice that a double application of the inversion operator amounts to identity: the argument of the function is unchanged. In other words, we have

$$\hat{P}^2\psi = P^2\psi = \psi, \quad \text{i.e.} \quad P^2 = 1, \quad \text{whence} \quad P = \pm 1. \qquad (19.3)$$

Thus the eigenfunctions of the inversion operator are either unchanged or change in sign when acted upon by this operator. In the first case, the wave function (and the corresponding state) is said to be *even*, and in the second it is said to be *odd*.

The invariance of the Hamiltonian under inversion (i.e. the fact that the operators $\hat{P}$ and $\hat{H}$ commute) thus expresses the *law of conservation of parity*: if the state of a closed system has a given parity (i.e. if it is even, or odd), then this parity is conserved in the course of time.‡

---

† Invariance under inversion exists also for the Hamiltonian of a system of particles in a centrally symmetric field with the centre at the origin.

‡ To avoid misunderstanding, it should be mentioned that this refers to the non-relativistic theory. There exist interactions in Nature, falling in the realm of relativistic theory, which violate the conservation of parity (see §90).

The angular momentum operator is also invariant under inversion, which changes the sign of the coordinates and of the operators of differentiation with respect to them; the operators (14.3) thus remain unaltered. In other words, the inversion operator commutes with the angular momentum operator, and this means that the system can have a definite parity simultaneously with definite values of the angular momentum L and its component M.

There are specific parity selection rules for the matrix elements of various physical quantities. Let us first consider scalars. Here we must distinguish *true scalars*, which are unchanged by inversion, from *pseudoscalars*, which change sign, for instance the scalar product of an axial and a polar vector.

It is easily seen that the matrix elements of a true scalar f can be different from zero only for transitions without change of parity: the matrix element for a transition between states of different parity is

$$f_{ug} = \int \psi_u^* \hat{f} \psi_g \, \mathrm{d}q,$$

where the function ψ_g is even and ψ_u odd, and the integrand changes sign when all the coordinates do so. But the integral over all space cannot be affected by merely renaming the variables of integration. Hence it follows that $f_{ug} = -f_{ug}$, or $f_{ug} = 0$. For a pseudoscalar, however, the matrix elements are different from zero only for transitions between states of different parity.

The selection rules for vector quantities can be obtained in a similar manner. Here it must be remembered that ordinary (*polar*) vectors change sign under inversion, but *axial* vectors are unchanged, for instance the angular momentum vector, which is the vector product of the two polar vectors **p** and **r**. Bearing this in mind, we find that the matrix elements are different from zero for transitions with change of parity for a polar vector, and without change of parity for an axial vector.

Let us determine the parity of the state of a single particle with angular momentum l. The inversion transformation ($x \to -x$, $y \to -y$, $z \to -z$) is, in spherical polar coordinates, the transformation

$$r \to r, \quad \theta \to \pi - \theta, \quad \phi \to \phi + \pi. \tag{19.4}$$

The dependence of the wave function of the particle on the angle is given by the eigenfunction Y_{lm} of the angular momentum (16.5). When ϕ is replaced by $\phi+\pi$, the factor $e^{im\phi}$ is multiplied by $(-1)^m$, and when θ is replaced by $\pi-\theta$, $P_l^m(\cos\theta)$ becomes $P_l^m(-\cos\theta) = (-1)^{l-m}P_l^m(\cos\theta)$. Thus the whole function is multiplied by $(-1)^l$, i.e. the parity of a state with a given value of l is

$$P = (-1)^l. \tag{19.5}$$

We see that all states with even l are even, and all those with odd l are odd. The parity of a state depends only on l, not on m.

Let us now determine the *addition rule for parities*. The wave function Ψ of a system consisting of two independent parts is the product of the wave functions Ψ_1 and Ψ_2 of these parts. Hence it is clear that, if the latter are of the same parity (i.e. both change sign, or both do not change sign, when the sign of all the coordinates is reversed), then the wave function of the whole system is even. On the other hand, if Ψ_1 and Ψ_2 are of opposite parity, then the function Ψ is odd.

This rule can be expressed by the equation

$$P = P_1P_2, \tag{19.6}$$

where P is the parity of the whole system, and P_1, P_2 those of its parts. It can, of course, be generalised at once to the case of a system composed of any number of non-interacting parts.

In particular, if we are concerned with a system of particles in a centrally symmetric field (the mutual interaction of the particles being supposed weak), the parity of the state of the whole system is

$$P = (-1)^{l_1+l_2+\cdots}. \tag{19.7}$$

We emphasise that the exponent here contains the algebraic sum of the angular momenta of the particles, and this is not in general the same as their "vector sum", i.e. the angular momentum L of the system.

If a closed system disintegrates (under the action of internal forces), the total angular momentum and parity must be conserved. This circumstance may render it impossible for a system to disintegrate, even if this is energetically possible.

For instance, let us consider an atom in an even state with angular momentum $L = 0$, which is able, so far as energy considerations go, to disintegrate into a free electron and an ion in an odd state with the same angular momentum $L = 0$. It is easy to see that in fact no such disintegration can occur (it is, as we say, *forbidden*). For by virtue of the law of conservation of angular momentum, the free electron would also have to have zero angular momentum, and therefore be in an even state ($P = (-1)^0 = 1$)); the state of the system ion + electron would then be odd, however, whereas the original state of the atom was even.

CHAPTER 3

SCHRÖDINGER'S EQUATION

§20. Schrödinger's equation

The form of the wave function of a physical system is determined by its Hamiltonian, which is therefore of fundamental significance in the whole mathematical formalism of quantum mechanics.

The form of the Hamiltonian for a free particle is established by the general requirements imposed by the homogeneity and isotropy of space and by Galileo's relativity principle. In classical mechanics, these requirements lead to a quadratic dependence of the energy of the particle on its momentum: $E = p^2/2m$, where the constant m is called the mass of the particle (see *Mechanics and Electrodynamics*, §4). In quantum mechanics, the same requirements lead to a corresponding relation for the energy and momentum eigenvalues, these quantities being conserved and simultaneously measurable (for a free particle).

If the relation $E = p^2/2m$ holds for every eigenvalue of the energy and momentum, the same relation must hold for their operators also:

$$\hat{H} = (1/2m)\,(\hat{p}_x^2+\hat{p}_y^2+\hat{p}_z^2). \tag{20.1}$$

Substituting here from (12.4), we obtain the Hamiltonian of a freely moving particle in the form

$$\hat{H} = -(\hbar^2/2m)\triangle, \tag{20.2}$$

where $\triangle = \partial^2/\partial x^2+\partial^2/\partial y^2+\partial^2/\partial z^2$ is the Laplacian operator.

If we have a system of non-interacting particles, its Hamiltonian is equal to the sum of the Hamiltonians of the separate particles:

$$\hat{H} = -\tfrac{1}{2}\hbar^2 \sum_a (1/m_a)\triangle_a \tag{20.3}$$

(the suffix a is the number of the particle; $\triangle_a$ is the Laplacian operator in which the differentiation is with respect to the coordinates of the ath particle).

In classical (non-relativistic) mechanics, the interaction of particles is described by an additive term in the Hamiltonian, the potential energy of the interaction $U(\mathbf{r}_1, \mathbf{r}_2, \ldots)$, which is a function of the coordinates of the particles. By adding a similar function to the Hamiltonian of the system, the interaction of particles can be represented in quantum mechanics:

$$\hat{H} = -\tfrac{1}{2}\hbar^2 \sum_a \triangle_a/m_a + U(\mathbf{r}_1, \mathbf{r}_2, \ldots). \tag{20.4}$$

The first term can be regarded as the operator of the kinetic energy and the second as that of the potential energy. The latter reduces to simple multiplication by the function U, and it follows from the passage to the limiting case of classical mechanics that this function must coincide with the one which gives the potential energy in classical mechanics. In particular, the Hamiltonian for a single particle in an external field is

$$\hat{H} = \hat{\mathbf{p}}^2/2m + U(x, y, z) = -(\hbar^2/2m)\triangle + U(x, y, z), \tag{20.5}$$

where $U(x, y, z)$ is the potential energy of the particle in the external field.

Substituting the expressions (20.2) to (20.5) in the general equation (8.1), we obtain the wave equations for the corresponding systems. We shall write out here the wave equation for a particle in an external field:

$$i\hbar\, \partial\Psi/\partial t = -(\hbar^2/2m)\triangle\Psi + U(x, y, z)\,\Psi. \tag{20.6}$$

The equation (10.2), which determines the stationary states, takes the form

$$(\hbar^2/2m)\triangle\psi + [E - U(x, y, z)]\,\psi = 0. \tag{20.7}$$

The equations (20.6) and (20.7) were obtained by Schrödinger in 1926 and are called *Schrödinger's equations.*

For a free particle, Schrödinger's equation (20.7) has the form

$$(\hbar^2/2m)\triangle\psi+E\psi = 0. \tag{20.8}$$

This equation has solutions finite in all space for any positive value of the energy E. For states with definite directions of motion, these solutions are eigenfunctions of the momentum operator (12.4), with $E = p^2/2m$. The complete (time-dependent) wave functions of such stationary states are

$$\Psi = \text{constant}\times e^{-(i/\hbar)(Et-\mathbf{p}.\mathbf{r})}. \tag{20.9}$$

Each such function is a *plane wave* and describes a state in which the particle has a definite energy E and momentum $\mathbf{p}$. This wave has a frequency $E/\hbar$ and wave vector $\mathbf{k} = \mathbf{p}/\hbar$ (the corresponding wavelength $\lambda = 2\pi\hbar/p$ is called the *de Broglie wavelength* of the particle).[†]

The energy spectrum of a freely moving particle is thus found to be continuous, extending from zero to $+\infty$. Each of these eigenvalues (except $E = 0$) is degenerate, and the degeneracy is infinite. For there corresponds to every value of E, different from zero, an infinite number of eigenfunctions (20.9), differing in the direction of the vector $\mathbf{p}$, which has a constant absolute magnitude.

§21. The current density

In classical mechanics, the velocity $\mathbf{v}$ of a particle is related to its momentum by $\mathbf{p} = m\mathbf{v}$. The same relation holds between the corresponding operators in quantum mechanics, as we should expect. This is easily shown by calculating the operator $\hat{\mathbf{v}} = \dot{\hat{\mathbf{r}}}$ by the general rule (9.2) for the differentiation of operators with respect to time.

[†] The idea of a wave related to a particle was first introduced by L. de Broglie in 1924.

Using the expression (20.5) for the Hamiltonian, we can write

$$\begin{aligned}\hat{\mathbf{v}} &= (i/\hbar)(\hat{H}\mathbf{r}-\mathbf{r}\hat{H})\\ &= -(i\hbar/2m)(\triangle\mathbf{r}-\mathbf{r}\triangle).\end{aligned}$$

To determine the value of the commutator, we apply it to an arbitrary function ψ:

$$\triangle(\mathbf{r}\psi)-\mathbf{r}(\triangle\psi) = 2\nabla\psi.$$

But $-i\hbar\nabla = \hat{\mathbf{p}}$, so that

$$\hat{\mathbf{v}} = \hat{\mathbf{p}}/m. \tag{21.1}$$

Similar relations will clearly hold between the eigenvalues of the velocity and momentum, and between their mean values in any state.

The velocity, like the momentum of a particle, cannot have a definite value simultaneously with the coordinates. But the velocity multiplied by an infinitely short time interval $\mathrm{d}t$ gives the displacement of the particle in the time $\mathrm{d}t$. Hence the fact that the velocity cannot exist at the same time as the coordinates means that, if the particle is at a definite point in space at some instant, it has no definite position at an infinitely close subsequent instant.

Next, let us find the acceleration operator. We have

$$\dot{\hat{\mathbf{v}}} = (i/\hbar)(\hat{H}\hat{\mathbf{v}}-\hat{\mathbf{v}}\hat{H}) = (i/m\hbar)(\hat{H}\hat{\mathbf{p}}-\hat{\mathbf{p}}\hat{H}) = (1/m)(U\nabla-\nabla U).$$

Here again, the operator can be evaluated by applying it to an arbitrary ψ:

$$U(\nabla\psi)-\nabla(U\psi) = -(\nabla U)\psi.$$

We therefore have

$$m\dot{\hat{\mathbf{v}}} = -\nabla U. \tag{21.2}$$

This operator equation is exactly the same in form as the equation of motion (Newton's equation) in classical mechanics.

The integral $\int|\Psi|^2\,\mathrm{d}V$, taken over some finite volume V, is the probability of finding the particle in this volume. Let us calculate the derivative of this probability with respect to time. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_V |\Psi|^2\,\mathrm{d}V = \int\limits_V \left(\Psi\frac{\partial\Psi^*}{\partial t}+\Psi^*\frac{\partial\Psi}{\partial t}\right)\mathrm{d}V$$

$$= \frac{i}{\hbar}\int\limits_V (\Psi\hat{H}^*\Psi^* - \Psi^*\hat{H}\Psi)\,\mathrm{d}V.$$

Substituting here

$$\hat{H} = \hat{H}^* = -(\hbar^2/2m)\triangle + U(x, y, z)$$

and using the identity

$$\Psi\triangle\Psi^* - \Psi^*\triangle\Psi = \operatorname{div}(\Psi\nabla\Psi^* - \Psi^*\nabla\Psi),$$

we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_V |\Psi|^2\,\mathrm{d}V = -\int\limits_V \operatorname{div}\mathbf{j}\,\mathrm{d}V,$$

where **j** denotes the vector

$$\mathbf{j} = (i\hbar/2m)(\Psi\nabla\Psi^* - \Psi^*\nabla\Psi) = \tfrac{1}{2}(\Psi^*\hat{\mathbf{v}}\Psi + \Psi\hat{\mathbf{v}}^*\Psi^*). \tag{21.3}$$

The integral of div **j** can be transformed by Gauss's theorem into an integral over the closed surface S which bounds[†] the volume V:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_V |\Psi|^2\,\mathrm{d}V = -\int\limits_S \mathbf{j}\cdot\mathrm{d}\mathbf{f}. \tag{21.4}$$

It is seen from this that the vector **j** may be called the *probability current density* vector or simply the *current density*. The integral of this vector over a surface is the probability that the particle will cross the surface during unit time. The vector **j** and the probability density $|\Psi|^2$ satisfy the equation

$$\partial|\Psi|^2/\partial t + \operatorname{div}\mathbf{j} = 0, \tag{21.5}$$

[†] The surface element **df** is, as usual, defined as a vector equal in magnitude to the area $\mathrm{d}f$ of the element and directed along the outward normal.

which is analogous to the classical equation of continuity (*Mechanics and Electrodynamics*, §55).

The wave function of free motion (the plane wave (20.9)) can be normalised so as to describe a flow of particles with unit current density (in which, on average, one particle crosses a unit cross-section of the flow per unit time). This function is then

$$\Psi = \frac{1}{\sqrt{v}} e^{(i/\hbar)(Et-\mathbf{p}.\mathbf{r})}, \tag{21.6}$$

where v is the velocity of the particle, since substitution of this in (21.3) gives $\mathbf{j} = \mathbf{p}/mv$, i.e. a unit vector in the direction of the motion.

§22. General properties of solutions of Schrödinger's equation

The conditions which must be satisfied by solutions of Schrödinger's equation are very general in character. First of all, the wave function and its first derivatives must be single-valued and continuous in all space. The requirement of continuity of the derivatives represents the condition for the current density to be continuous.

If the field $U(x, y, z)$ nowhere becomes infinite, then the wave function also must be finite in all space. The same condition must hold in cases where U becomes infinite at some point but does not do so too rapidly.[†]

Let $U_{\min}$ be the least value of the function $U(x, y, z)$. Since the Hamiltonian is the sum of two terms, the operators of the kinetic energy $(\hat{T})$ and of the potential energy, the mean value $\bar{E}$ of the energy in any state is equal to the sum $\bar{T}+\bar{U}$. But all the eigenvalues of the operator $\hat{T}$ (which is the Hamiltonian of a free particle) are positive; hence the mean value $\bar{T} \geqslant 0$. Recalling also the obvious inequality $\bar{U} > U_{\min}$, we find that $\bar{E} > U_{\min}$. Since this inequality holds for any

---

[†] To wit, not more rapidly than $-1/r^2$, where r is the distance from the point. It can be shown that if U tends to $-\infty$ more rapidly than this, the "normal" state will correspond to a particle at the point $r = 0$, i.e. the particle "falls" to this point.

state, it is clear that it is valid for all the eigenvalues of the energy:

$$E_n > U_{\min}. \tag{22.1}$$

Let us consider a particle moving in an external field which vanishes at infinity; we define the function $U(x, y, z)$, in the usual way, so that it vanishes at infinity. It is easy to see that the spectrum of negative eigenvalues of the energy will then be discrete, i.e. all states with $E < 0$ are bound states. For, in the stationary states of a continuous spectrum, which correspond to infinite motion, the particle reaches infinity (see §10); however, at sufficiently large distances the field may be neglected, the motion of the particle may be regarded as free, and the energy of a freely moving particle can only be positive.

The positive eigenvalues, on the other hand, form a continuous spectrum and correspond to an infinite motion; for $E > 0$, Schrödinger's equation in general has no solutions (in the field concerned) for which the integral $\int |\psi|^2 \, dV$ converges.

In quantum mechanics, a particle in a finite motion may be found in those regions of space where $E < U$; the probability $|\psi|^2$ of finding the particle tends rapidly to zero as the distance into such a region increases, yet it differs from zero at all finite distances. Here there is a fundamental difference from classical mechanics, in which a particle cannot penetrate into a region where $U > E$. In classical mechanics the impossibility of penetrating into this region is related to the fact that, for $E < U$, the kinetic energy would be negative, that is, the velocity would be imaginary, which is meaningless. In quantum mechanics, the eigenvalues of the kinetic energy are likewise positive; nevertheless, we do not reach a contradiction here, since, if by a process of measurement a particle is localised at some definite point of space, the state of the particle is changed, as a result of this process, in such a way that it ceases in general to have any definite kinetic energy.

The above discussion may be illustrated by examples of one-dimensional motion, i.e. motion in a field $U(x)$ that depends on only one coordinate. The motion in the y- and z-directions is then free, while that along the x-axis is governed by the one-dimensional

Schrödinger's equation

$$\frac{d^2\psi}{dx^2}+\frac{2m}{\hbar^2}[E-U(x)]\psi = 0. \tag{22.2}$$

In the "potential well" shown in Fig. 1a, the motion with energy $E < 0$ is finite, and the corresponding energy spectrum is discrete.

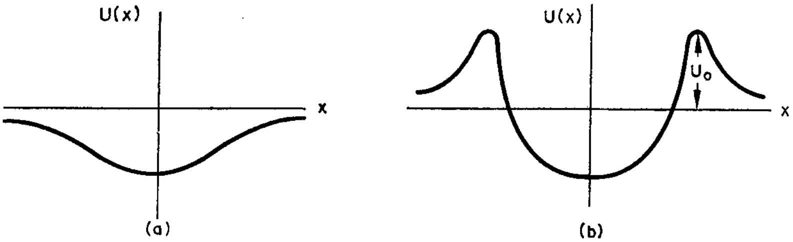

FIG. 1

The energies $E > 0$, however, have a continuous spectrum, and the motion is infinite. Let us determine the asymptotic form of the wave functions at large distances x in these two cases. Since $U \to 0$ when $x \to \pm\infty$, the field U can be neglected in comparison with E in equation (22.2), so that

$$\frac{d^2\psi}{dx^2}+\frac{2m}{\hbar^2}E\psi = 0. \tag{22.3}$$

When $E > 0$, this is the equation of free motion in one dimension; its general solution is

$$\psi = a_1e^{ikx}+a_2e^{-ikx}, \quad k = (1/\hbar)\sqrt{(2mE)}, \tag{22.4}$$

i.e. is a superposition of two plane waves corresponding to motion to the right and to the left along the x-axis. Each energy level is doubly degenerate in accordance with the two possible motions in opposite directions.

For energies $E < 0$, only one of the two independent solutions of the second-order differential equation (22.2) is permissible, since it must satisfy the boundary conditions whereby the wave function for a finite motion must tend to zero as $x \to \pm\infty$. At large distances, we again have equation (22.3), but its solution has the asymptotic form

$$\psi = \text{constant}\times e^{\mp\varkappa x} \text{ as } x \to \pm\infty, \quad \varkappa = (1/\hbar)\sqrt{(2m\,|E|)}, \tag{22.5}$$

i.e. is exponentially damped in the classically inaccessible region (the other solution of (22.3) increases without limit as $x \to \pm\infty$).

So far as finite and infinite motion alone are concerned, both occur in the corresponding cases in classical and quantum mechanics ($E < 0$ and $E > 0$ respectively) for a field of the type shown in Fig. 1a. This is no longer true, however, for the field shown in Fig. 1b, where the well is surrounded by a "potential barrier" of finite height U_0. The motion with $E < 0$ is again finite. In classical mechanics it would also be finite for motion within the well with $0 < E < U_0$. In quantum mechanics, however, the motion is infinite for any energy $E > 0$, whether greater than or less than the height of the potential barrier. A particle (with $E > 0$) that is "inside the well" at a certain instant may later pass "though the barrier" and reach the region outside the well.

Thus quantum mechanics allows infinite motion of particles under conditions where this could not occur in classical mechanics. The nature of this passage through the barrier (which will be further discussed in §28) is related to the above-mentioned fact that the wave function is not exactly zero within the classically inaccessible region.

Schrödinger's equation, in the general form $\hat{H}\psi = E\psi$, can be obtained from the variational principle

$$\delta \int \psi^*(\hat{H}-E)\psi\, \mathrm{d}q = 0. \tag{22.6}$$

Since ψ is complex, we can vary ψ and ψ^* independently. Varying ψ^*, we have

$$\int \delta\psi^*(\hat{H}-E)\psi\, \mathrm{d}q = 0,$$

whence, because $\delta\psi^*$ is arbitrary, we obtain the required equation $\hat{H}\psi = E\psi$. The variation of ψ gives nothing different; we obtain only the complex conjugate equation $\hat{H}^*\psi^* = E\psi^*$.

The variational calculus can be used to prove several important results concerning the general properties of the wave functions of steady states of a particle.

The wave function ψ_0 of the normal state does not become zero (or, as we say, has no *nodes*) for any finite values of the coordinates. In other words, it has the same sign in all space. Hence, it follows that the wave functions ψ_n $(n > 0)$ of the other stationary states, being orthogonal to ψ_0, must have nodes (if ψ_n is also of constant sign, the integral $\int \psi_0 \psi_n \, \mathrm{d}V$ cannot vanish).

Next, from the fact that ψ_0 has no nodes, it follows that the normal energy level cannot be degenerate. For, suppose the contrary to be true, and let ψ_0, ψ_0' be two different eigenfunctions corresponding to the level E_0. Any linear combination $c\psi_0 + c'\psi_0'$ will also be an eigenfunction; but by choosing the appropriate constants c, c', we can always make this function vanish at any given point in space, i.e. we can obtain an eigenfunction with nodes.

For a one-dimensional motion we have the more restrictive *oscillation theorem*: the wave function $\psi_n(x)$ of a discrete spectrum corresponding to the $(n+1)$th eigenvalue E_n (the eigenvalues being arranged in order of magnitude) vanishes n times (for finite values of x).

§23. **Time reversal**

Schrödinger's equation for the wave functions of stationary states is real, as are the conditions imposed on its solution. Hence its solutions ψ can always be taken as real. The eigenfunctions of non-degenerate values of the energy are automatically real, apart from the unimportant phase factor. For ψ^* satisfies the same equation as ψ, and therefore must also be an eigenfunction for the same value of the energy; hence, if this value is not degenerate, ψ and ψ^* must be essentially the same, i.e. they can differ only by a constant phase factor. The wave functions corresponding to the same degenerate energy level need not be real, however, but by a suitable choice of linear combinations of them we can always obtain a set of real functions.

The complete (time-dependent) wave functions Ψ are determined by an equation in whose coefficients i appears. This equation, however, retains the same form if we replace t in it by $-t$ and at the same time take the complex conjugate. Hence we can always choose the functions Ψ in such a way that Ψ and Ψ^* differ only by the sign of the time, a result which we know already from formulae (10.1) and (10.3).

As is well known, the equations of classical mechanics are unchanged by *time reversal*, i.e. when the sign of the time is reversed. In quantum mechanics, the symmetry with respect to the two directions of time is expressed, as we see, in the invariance of the wave equation when the sign of t is changed and Ψ is simultaneously replaced by Ψ^*. However, it must be recalled that this symmetry here relates only to the wave equation, and not to the concept of measurement itself, which plays a fundamental part in quantum mechanics. The measuring process in quantum mechanics has a "two-faced" character: it plays different parts with respect to the past and the future. With respect to the past, it "verifies" the probabilities of the various possible results predicted from the state brought about by the previous measurement. With respect to the future, it brings about a new state (see also §37). Thus the very nature of the quantum-mechanical process of measurement involves a far-reaching principle of irreversibility.

This irreversibility is of fundamental significance. Although the basic equations of quantum mechanics are in themselves symmetrical with respect to a change in the sign of the time (in this respect quantum mechanics does not differ from classical mechanics), the irreversibility of the process of measurement causes the two directions of time to be physically non-equivalent, i.e. creates a difference between the future and the past.

§24. The potential well

As a simple example of one-dimensional motion, let us consider motion in a square potential well as shown in Fig. 2 (it will here be more convenient to reckon the energy from the bottom of the well, not from the value of the potential energy at infinity). We shall inves-

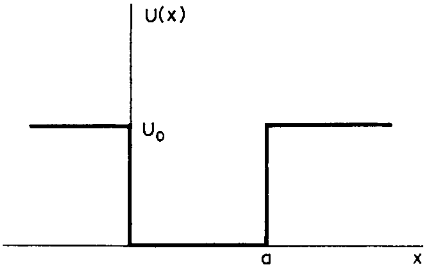

FIG. 2

tigate the states of finite motion belonging to the discrete energy spectrum $0 < E < U_0$.

In the region $0 < x < a$ we have Schrödinger's equation

$$\psi''+k^2\psi = 0, \quad k = (1/\hbar)\sqrt{(2mE)} \tag{24.1}$$

(the prime denotes differentiation with respect to x), while in the region outside the well

$$\psi''-\varkappa^2\psi = 0, \quad \varkappa = (1/\hbar)\sqrt{[2m(U_0-E)]}. \tag{24.2}$$

For $x = 0$ and $x = a$ the solutions of these equations must be continuous together with their derivatives.

The solution of equation (24.2) which vanishes at infinity is

$$\psi = \text{constant}\times e^{\mp\varkappa x}; \tag{24.3}$$

the signs $-$ and $+$ in the exponent refer to the regions $x > a$ and $x < 0$ respectively. Instead of the continuity of ψ and ψ' at the edge of the potential well, it is convenient to require the continuity of ψ and of its logarithmic derivative ψ'/ψ. Taking account of (24.3), we obtain the boundary condition in the form

$$\psi'/\psi = \mp\varkappa. \tag{24.4}$$

We shall not pause here to determine the energy levels in a well of arbitrary depth U_0 (see Problem 2), and shall analyse fully only the limiting case of infinitely high walls.

When $U_0 \to \infty$, the function (24.3) is identically zero: the particle cannot, of course, reach a region where the potential is infinite. Thus we have to find the solution of equation (24.1) with the boundary condition

$$\psi = 0 \quad \text{for} \quad x = 0 \text{ and } a. \tag{24.5}$$

Such a solution may be sought as a "stationary wave"

$$\psi = c \sin (kx+\delta). \tag{24.6}$$

The condition $\psi = 0$ for $x = 0$ gives $\delta = 0$, and then the condition at $x = a$ gives $\sin ka = 0$, whence $ka = (n+1)\pi$, with $n = 0, 1, 2, \ldots$.

The energy levels of a particle in the well are therefore

$$E_n = (\pi^2\hbar^2/2ma^2)(n+1)^2, \qquad n = 0, 1, 2, \ldots . \tag{24.7}$$

In particular, the energy of the ground state is $E_0 = \pi^2\hbar^2/2ma^2$. This result is in accordance with the uncertainty relation: when the uncertainty of the coordinate is proportional to a, that of the momentum (and therefore the order of magnitude of the momentum itself) is proportional to $\hbar/a$, and the corresponding energy is proportional to $(\hbar/a)^2m$.

The normalised wave functions of the stationary states are

$$\psi_n = \sqrt{\frac{2}{a}} \sin \frac{\pi(n+1)x}{a}. \tag{24.8}$$

In accordance with the oscillation theorem, the function $\psi_n(x)$ is zero at n points within the region of the motion (the boundaries of the region, in this case the points $x = 0$ and a, are excluded from the zeros in applying the oscillation theorem).

In a one-dimensional potential well of any shape, there is always at least one energy level, even if the well is very shallow (see, for instance, Problem 2). This is, however, a specific property of the one-dimensional case, and does not occur in the more realistic three-dimensional well; if the depth $|U|$ of such a well is

$$|U| \ll \hbar^2/ma^2, \tag{24.9}$$

where a is the order of magnitude of the linear dimensions of the well, it has no discrete energy levels. Thus, if the well is not sufficiently deep, it has no bound states, and the particle cannot be "captured" by the well. This is a purely quantum property; in classical mechanics, a particle can execute a finite motion in any potential well. The reason for the property will be explained in §32, and it will be proved by direct calculation for the case of a spherically symmetric well in §30, Problem 1.

PROBLEMS

PROBLEM 1. Determine the probability distribution for various values of the momentum for the normal state of a particle in one-dimensional motion in an infinitely deep square potential well.

SOLUTION. The probability of momentum values p lying in the range $\mathrm{d}p$ is $|a(p)|^2\,\mathrm{d}p$, where $a(p)$ in the one-dimensional case is

$$a(p) = \frac{1}{\sqrt{(2\pi\hbar)}}\int\limits_0^a \psi_0(x)e^{-(i/\hbar)px}\,\mathrm{d}x;$$

cf. (12.12). Substituting $\psi_0(x)$ from (24.8) and calculating the integral, we obtain the required probability distribution:

$$|a(p)|^2 = \frac{4\pi\hbar^3 a}{(p^2a^2-\pi^2\hbar^2)^2}\cos^2\frac{pa}{2\hbar}.$$

PROBLEM 2. Determine the energy levels for the potential well shown in Fig. 2.

SOLUTION. The condition (24.4) at the edges of the well gives the equations

$$k\cot\delta = -k\cot(ka+\delta) = \varkappa \equiv \sqrt{[(2m/\hbar^2)U_0-k^2]},$$

or

$$\sin\delta = -\sin(ka+\delta) = k\hbar/\sqrt{(2mU_0)}.$$

Eliminating δ, we obtain the transcendental equation

$$ka = (n+1)\pi - 2\sin^{-1}[k\hbar/\sqrt{(2mU_0)}], \tag{1}$$

where $n = 0, 1, 2, \ldots$, and the values of the inverse sine are taken between 0 and $\frac{1}{2}\pi$. The roots of this equation determine the energy levels $E = k^2\hbar^2/2m$. The values of n number the levels in order of increasing energy. The number of levels is finite if U_0 is finite.

Equation (1) can be written in a more convenient form by using the variable ξ and the parameter γ defined by

$$\xi = \tfrac{1}{2}ka, \quad \gamma = (\hbar/a)\sqrt{(2/mU_0)}.$$

When n is even, the resulting equation is

$$\cos\xi = \pm\gamma\xi, \tag{2}$$

and the roots for which $\tan\xi > 0$ must be taken. When n is odd, we have

$$\sin\xi = \pm\gamma\xi, \tag{3}$$

and the roots for which $\tan\xi < 0$ must be taken.

In particular, for a shallow well in which $U_0 \ll \hbar^2/ma^2$, we have $\gamma \gg 1$ and equation (3) has no root. Equation (2) has one root (with the upper sign on the right-hand side), $\xi \cong 1/\gamma - 1/2\gamma^3$. Thus the well contains only one energy level,

$$E_0 = 2\xi^2\hbar^2/ma^2 \cong U_0 - (ma^2/2\hbar^2)U_0^2,$$

which is near the top of the well.

Problem 3. Determine the energy levels of a particle moving in a rectangular "potential box" with sides a, b, c : $U = 0$ inside the box and $U = \infty$ outside it.

Solution. The free motion of the particle within the box takes place independently in three directions. The energy levels are therefore given simply by the sums of three expressions like (24.7):

$$E_{n_1n_2n_3} = \frac{\pi^2\hbar^2}{2m}\left(\frac{n_1^2}{a^2}+\frac{n_2^2}{b^2}+\frac{n_3^2}{c^2}\right) \qquad (n_1, n_2, n_3 = 1, 2, \ldots).$$

The intervals between the levels tend to zero as the size of the well increases. The wave functions of the stationary states are

$$\psi_{n_1n_2n_3} = \sqrt{\left(\frac{8}{abc}\right)}\sin\frac{\pi n_1 x}{a}\sin\frac{\pi n_2 y}{b}\sin\frac{\pi n_3 z}{c},$$

where the axes of x, y, z are along the sides of the box.

§25. The linear oscillator

Let us consider a particle executing small oscillations in one dimension (what is called a *linear oscillator*). The potential energy of such a particle is $\frac{1}{2}m\omega^2x^2$, where ω is, in classical mechanics, the characteristic (angular) frequency of the oscillations (see *Mechanics and Electrodynamics*, §17). Accordingly, the Hamiltonian of the oscillator is

$$\hat{H} = \tfrac{1}{2}\hat{p}^2/m + \tfrac{1}{2}m\omega^2x^2. \tag{25.1}$$

Since the potential energy becomes infinite for $x = \pm\infty$, the particle can have only a finite motion, and the energy eigenvalue spectrum is entirely discrete.

Let us determine the energy levels of the oscillator, using the matrix method.[†] We shall start from the "equations of motion" in the form

---

[†] This was done by Heisenberg in 1925, before Schrödinger's discovery of the wave equation.

(21.2); in this case they give

$$\hat{\ddot{x}}+\omega^2 x = 0. \tag{25.2}$$

In matrix form, this equation reads

$$(\ddot{x})_{mn}+\omega^2 x_{mn} = 0.$$

For the matrix elements of the acceleration we have, according to (11.8), $(\ddot{x})_{mn} = i\omega_{mn}(\dot{x})_{mn} = -\omega_{mn}^2 x_{mn}$. Hence we obtain

$$(\omega_{mn}^2-\omega^2)x_{mn} = 0.$$

Hence it is evident that all the matrix elements x_{mn} vanish except those for which $\omega_{mn} = \omega$ or $\omega_{mn} = -\omega$. We number all the stationary states so that the frequencies $\pm\omega$ correspond to transitions $n \rightarrow n\mp 1$, i.e. $\omega_{n,\,n\mp1} = \pm\omega$. Then the only non-zero matrix elements are $x_{n,\,n\pm1}$.

We shall suppose that the wave functions ψ_n are taken real. Since x is a real quantity, all the matrix elements x_{mn} are real. The Hermitian condition (11.10) now shows that the matrix x_{mn} is symmetrical:

$$x_{mn} = x_{nm}.$$

To calculate the matrix elements of the coordinate which are different from zero, we use the commutation rule

$$\hat{\dot{x}}\hat{x}-\hat{x}\hat{\dot{x}} = -i\hbar/m,$$

written in the matrix form

$$(\dot{x}x)_{mn}-(x\dot{x})_{mn} = -(i\hbar/m)\delta_{mn}.$$

By the matrix multiplication rule (11.12) we hence have for $m = n$

$$i\sum_l(\omega_{nl}x_{nl}x_{ln}-x_{nl}\omega_{ln}x_{ln}) = 2i\sum_l\omega_{nl}x_{nl}^2 = -i\hbar/m.$$

In this sum, only the terms with $l = n\pm1$ are different from zero, so that we have

$$(x_{n+1,\,n})^2-(x_{n,\,n-1})^2 = \hbar/2m\omega. \tag{25.3}$$

From this equation we deduce that the quantities $(x_{n+1,\,n})^2$ form an arithmetic progression, which is unbounded above, but is certainly bounded below, since it can contain only positive terms. Since we have as yet fixed only the relative positions of the numbers n of the states, but not their absolute values, we can arbitrarily choose the value of n corresponding to the first (normal) state of the oscillator, and put this value equal to zero. Accordingly $x_{0,\,-1}$ must be regarded as being zero identically, and the application of equations (25.3) with $n = 0, 1, \ldots$ successively leads to the result

$$(x_{n,\,n-1})^2 = n\hbar/2m\omega.$$

Thus we finally obtain the following expression for the matrix elements of the coordinate which are different from zero:

$$x_{n,\,n-1} = x_{n-1,\,n} = \sqrt{(n\hbar/2m\omega)}. \tag{25.4}$$

The matrix of the operator $\hat{H}$ is diagonal, and the matrix elements H_{nn} are the required eigenvalues E_n of the energy of the oscillator. To calculate them, we write

$$\begin{aligned} H_{nn} = E_n &= \tfrac{1}{2}m[(\dot{x}^2)_{nn}+\omega^2(x^2)_{nn}] \\ &= \tfrac{1}{2}m\Big[\sum_l i\omega_{nl}x_{nl}i\omega_{ln}x_{ln}+\omega^2\sum_l x_{nl}x_{ln}\Big] \\ &= \tfrac{1}{2}m\sum_l(\omega^2+\omega_{nl}^2)x_{ln}^2. \end{aligned}$$

In the sum over l, only the terms with $l = n\pm 1$ are different from zero; substituting (25.4), we obtain

$$E_n = (n+\tfrac{1}{2})\hbar\omega, \qquad n = 0, 1, 2, \ldots\ . \tag{25.5}$$

Thus the energy levels of the oscillator lie at equal intervals of $\hbar\omega$ from one another. The energy of the normal state ($n = 0$) is $\frac{1}{2}\hbar\omega$; we call attention to the fact that it is not zero.

The result (25.5) can also be obtained by solving Schrödinger's equation. For an oscillator, this has the form

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+\frac{2m}{\hbar^2}\left(E-\tfrac{1}{2}m\omega^2x^2\right)\psi = 0. \tag{25.6}$$

Here it is convenient to introduce, instead of the coordinate x, the dimensionless variable ξ by the relation

$$\xi = \sqrt{(m\omega/\hbar)}x. \tag{25.7}$$

Then we have the equation

$$\psi'' + [(2E/\hbar\omega) - \xi^2]\psi = 0; \tag{25.8}$$

here the prime denotes differentiation with respect to ξ.

For large ξ, we can neglect $2E/\hbar\omega$ in comparison with ξ^2; the equation $\psi'' = \xi^2\psi$ has the asymptotic integrals $\psi = e^{\pm\frac{1}{2}\xi^2}$ (for differentiation of this function gives $\psi'' = \xi^2\psi$ on neglecting terms of order less than that of the term retained). Since the wave function ψ must remain finite as $\xi \to \pm\infty$, the index must be taken with the minus sign. It is therefore natural to make in equation (25.8) the substitution

$$\psi = e^{-\xi^2/2}\chi(\xi). \tag{25.9}$$

For the function $\chi(\xi)$ we obtain the equation (with the notation $(2E/\hbar\omega) - 1 = 2n$)

$$\chi'' - 2\xi\chi' + 2n\chi = 0; \tag{25.10}$$

the function χ must be finite for all finite ξ, and for $\xi \to \pm\infty$ must not tend to infinity more rapidly than every finite power of ξ (in order that the function ψ should tend to zero).

We shall seek a solution of equation (25.10) as a series

$$\chi = \sum_{s=0}^{\infty} a_s \xi^s. \tag{25.11}$$

Substitution of this gives

$$\sum_{s=2}^{\infty} a_s s(s-1)\xi^{s-2} - 2\sum_{s=0}^{\infty} a_s s \xi^s + 2n \sum_{s=0}^{\infty} a_s \xi^s = 0.$$

In the first sum, we rename the variable of summation, replacing s by $s+2$:

$$\sum_{s=0}^{\infty} [a_{s+2}(s+1)(s+2) + 2(n-s)a_s]\xi^s = 0.$$

If this is identically satisfied, the coefficient of each power of ξ must be zero. Thus we find the recurrence relation

$$a_{s+2} = -\frac{2(n-s)}{(s+1)(s+2)} a_s \tag{25.12}$$

between the coefficients of successive terms in the series (25.11). It is seen, first of all, that the series contains only odd or only even powers of ξ. In order to satisfy the condition stated above, the series must contain only terms with finite powers, i.e. it must stop at some finite s. From (25.12), n must therefore be a positive integer; the series then stops at $s = n$, becoming a polynomial of degree n. This is the result (25.5) already obtained for the energy eigenvalues.

We shall give the explicit form of the wave function only for the ground state of the oscillator. When $n = 0$, the polynomial reduces to a constant. Determining this so that the wave function satisfies the normalisation condition

$$\int\limits_{-\infty}^{\infty} \psi_0^2(x)\, \mathrm{d}x = 1,$$

we find

$$\psi_0(x) = (m\omega/\pi\hbar)^{1/4} e^{-m\omega x^2/2\hbar}. \tag{25.13}$$

This function has no zeros for finite x, which is as it should be.

PROBLEM

Determine the probability distribution of the various values of the momentum, in the normal state of the oscillator.

SOLUTION. As in §24, Problem 1, we calculate the integral

$$a(p) = \frac{1}{\sqrt{(2\pi\hbar)}} \int\limits_{-\infty}^{\infty} \psi_0(x) e^{-(i/\hbar)px}\, \mathrm{d}x.$$

The substitution $x + ip/m\omega = z$ reduces this to a Poisson integral, and the result is

$$|a(p)|^2 = \frac{1}{\sqrt{(\pi m\hbar\omega)}} e^{(-p^2/m\hbar\omega)}.$$

§26. The quasi-classical wave function

If the de Broglie wavelengths of particles are small in comparison with the characteristic dimensions which determine the conditions of a given problem, then the properties of the system are close to being classical. In §6 we have already mentioned the general form of the wave functions in such quasi-classical cases, and in §§12 and 14 this form has been used to derive the quantum-mechanical operators of fundamental physical quantities. We shall now investigate more closely how the passage to the quasi-classical limit takes place in Schrödinger's equation.

It has been noted in §6 that the transition from quantum mechanics to classical mechanics can be formally described as a passage to the limit $\hbar \to 0$. In the quasi-classical case, therefore, $\hbar$ may be regarded as a small parameter, and the expression

$$\Psi = ae^{iS/\hbar}, \tag{26.1}$$

in which a and S are assumed independent of $\hbar$, can be regarded as the first term in an expansion of the wave function in powers of that parameter. If (26.1) is put in the form $\exp\{(iS+\hbar\log a)/\hbar\}$, we see that it corresponds to the first two terms in the expansion of the exponential. In the subsequent calculations, therefore, only the first two powers of $\hbar$ need be retained.

For simplicity, we shall refer to a single particle in an external field. Substituting (26.1) in Schrödinger's equation (20.6), differentiating and retaining only the first two powers of $\hbar$, we get

$$a\frac{\partial S}{\partial t} - i\hbar\frac{\partial a}{\partial t} + \frac{a}{2m}(\nabla S)^2 - \frac{i\hbar}{2m}a\triangle S - \frac{i\hbar}{m}\nabla S.\nabla a + Ua = 0. \tag{26.2}$$

Equating the powers of $\hbar$ separately to zero, we obtain two equations:

$$\frac{\partial S}{\partial t} + \frac{1}{2m}(\nabla S)^2 + U = 0, \tag{26.3}$$

$$\frac{\partial a}{\partial t} + \frac{a}{2m}\triangle S + \frac{1}{m}\nabla S.\nabla a = 0. \tag{26.4}$$

The first of these is the Hamilton–Jacobi equation for the action S of a particle, as it should be (see *Mechanics and Electrodynamics*, §31). Equation (26.4), on multiplication by $2a$, can be rewritten in the form

$$\frac{\partial a^2}{\partial t}+\operatorname{div}\left(a^2\frac{\nabla S}{m}\right)=0. \tag{26.5}$$

This equation has an obvious physical meaning: $|\Psi|^2=a^2$ is the probability density for finding the particle at some point in space; $\nabla S/m=\mathbf{p}/m$ is the classical velocity $\mathbf{v}$ of the particle. Hence equation (26.5) is simply the equation of continuity, which shows that the probability density "moves" according to the laws of classical mechanics with the classical velocity $\mathbf{v}$ at every point.

For stationary states, i.e. for a given energy E, the action is

$$S=-Et+S_0(x,y,z), \tag{26.6}$$

where S_0 is a function of the coordinates (called the "abbreviated action") and satisfies the equation

$$\frac{1}{2m}(\nabla S_0)^2+U=E. \tag{26.7}$$

The amplitude a of the wave function for the stationary states is independent of time, and satisfies the equation

$$\operatorname{div}(a^2\nabla S)=0. \tag{26.8}$$

The quasi-classical function for stationary states will be given explicitly for one-dimensional motion of a particle in a field $U(x)$. Then, in equation (26.7), $(\nabla S_0)^2=(\mathrm{d}S_0/\mathrm{d}x)^2$, and the solution is

$$S_0=\pm\int p\,\mathrm{d}x,\quad p(x)=\sqrt{[2m(E-U)]}. \tag{26.9}$$

The integrand $p(x)$ is just the classical momentum of the particle, expressed as a function of the coordinate. From (26.8) we then have

$$\mathrm{d}(a^2p)/\mathrm{d}x=0,\quad a^2p=\text{constant},$$

so that $a=\text{constant}/\sqrt{p}$. Thus the general solution of Schrödinger's equation is

$$\psi = \frac{C_1}{\sqrt{p}} e^{(i/\hbar)\int p\,\mathrm{d}x} + \frac{C_2}{\sqrt{p}} e^{-(i/\hbar)\int p\,\mathrm{d}x}, \tag{26.10}$$

where C_1 and C_2 are constant coefficients.

The presence of the factor $1/\sqrt{p}$ in the wave function has a simple explanation. The probability of finding the particle at a point with coordinate between x and $x+\mathrm{d}x$ is determined by $|\psi|^2$, and is therefore essentially proportional to $1/p$. This is as we should expect for a "quasi-classical particle", since in classical motion the time spent by the particle in the segment $\mathrm{d}x$ is inversely proportional to its velocity (or momentum).

In the "classically inaccessible" regions of space, where $E < U(x)$, the function $p(x)$ is purely imaginary, so that the exponents are real. The wave function in these regions can be written in the form

$$\psi = \frac{C_1'}{\sqrt{|p|}} e^{-(1/\hbar)\int |p|\,\mathrm{d}x} + \frac{C_2'}{\sqrt{|p|}} e^{(1/\hbar)\int |p|\,\mathrm{d}x}. \tag{26.11}$$

Let us examine more closely the condition for these results to be valid. In equation (26.2) the terms in $\hbar$ must in fact be small in comparison with the others. We may compare, for example, the terms

$$\frac{a}{2m}(\nabla S)^2 = \frac{a}{2m}\left(\frac{\mathrm{d}S}{\mathrm{d}x}\right)^2 = \frac{a}{2m}p^2,$$

$$\frac{i\hbar a}{2m}\triangle S = \frac{i\hbar a}{2m}\frac{\mathrm{d}^2S}{\mathrm{d}x^2} = \frac{i\hbar a}{2m}\frac{\mathrm{d}p}{\mathrm{d}x}.$$

The condition for the latter to be small in comparison with the former is $(\hbar/p^2)\,|\mathrm{d}p/\mathrm{d}x| \ll 1$, or

$$|\mathrm{d}ƛ/\mathrm{d}x| \ll 1, \tag{26.12}$$

where $ƛ = \lambda/2\pi$ and $\lambda(x) = 2\pi\hbar/p(x)$ is the de Broglie wavelength of the particle, expressed as a function of x by means of the classical function $p(x)$. Thus we have a quantitative criterion of quasi-classicality: the wavelength of the particle must vary only slightly over a distance of the order of this wavelength. The formulae derived above become inapplicable in regions of space where this condition is not satisfied.

The quasi-classical approximation is clearly inapplicable near *turning points*, i.e. near points where the particle, according to classical mechanics, would stop and begin to move in the opposite direction. These points are given by the equation $p(x) = 0$. As $p \to 0$, the de Broglie wavelength tends to infinity, and hence cannot possibly be supposed small.

§27. Bohr and Sommerfeld's quantisation rule

The results which we have obtained in §26 enable us to derive the condition which determines the quantum energy levels in the quasi-classical case. To do this we consider a finite one-dimensional motion of a particle in a potential well: the classically accessible region $a \leq x \leq b$ is bounded by two turning points (Fig. 3).[†]

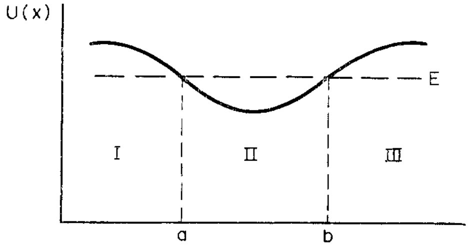

FIG. 3

The boundary conditions for the wave function consist in the requirement that it is damped in each of the classically inaccessible re-

---

[†] In classical mechanics, a particle in such a field would execute a periodic motion with period (time taken in moving from $x = a$ to $x = b$ and back)

$$T = 2\int_a^b dx/v = 2m\int_a^b dx/p,$$

where v is the velocity of the particle.

gions I and III, becoming zero as $x \to \pm\infty$. We know also that in these regions the general solution of Schrödinger's equation has the form (26.11), and in region II it is (26.10). From these conditions it would be possible to determine the constant coefficients in the solution for each region by joining them at the boundaries $x = a$ and $x = b$. But this joining cannot be achieved directly, because near such points the quasi-classical approximation used to calculate (26.10) and (26.11) becomes invalid.

The difficulty is eliminated if we use only the crude approximation of making the wave function vanish at $x = a$ and $x = b$, not at infinity.

In the classical limit, these points are the absolute limits of the motion, and the particle cannot pass beyond them. In the quasi-classical approximation, although the particle can penetrate into classically inaccessible regions, the wave functions are damped very rapidly there; this is the basis for the change of boundary conditions mentioned in the previous paragraph.

The boundary condition $\psi = 0$ for $x = a$ gives for the wave function in region II the expression

$$\psi = \frac{C}{\sqrt{p}} \sin \frac{1}{\hbar} \int_a^x p \, \mathrm{d}x. \tag{27.1}$$

Similarly, substituting the condition $\psi = 0$ at $x = b$, we have

$$\psi = \frac{C'}{\sqrt{p}} \sin \frac{1}{\hbar} \int_x^b p \, \mathrm{d}x.$$

If these two expressions are the same throughout the region, the sum of their phases (which is a constant) must be an integral multiple of π:

$$\frac{1}{\hbar} \int_a^b p \, \mathrm{d}x = n\pi, \tag{27.2}$$

with $C = (-1)^n C'$. This may be also written

$$\oint p \, \mathrm{d}x = 2\pi\hbar n, \tag{27.3}$$

where the integral is taken over the whole period of the classical motion of the particle. This is the condition which determines the stationary states of the particle in the quasi-classical case. It corresponds to Bohr and Sommerfeld's quantisation rule in the old quantum theory.

Since $\hbar$ is a small parameter in the quasi-classical approximation, the expression on the left of equation (27.2) is large. The same is therefore true of the integer n. The phase of the wave function (27.1) varies from 0 at $x = a$ to $n\pi$ at $x = b$, and the sine therefore vanishes $n-1$, or approximately n, times in this interval. Thus the integer n represents the number of zeros of the wave function. According to the oscillation theorem (§22) it is therefore a quantum number that identifies the successive quantum energy levels.[†]

The fact that the quasi-classical approximation corresponds to a large value of the quantum number n has a simple intuitive explanation. The distance between adjacent zeros of the wave function is evidently equal, in order of magnitude, to the de Broglie wavelength. For large n, this distance is small, being approximately $(b-a)/n$; the wavelength is therefore small in comparison with the dimensions of the region of motion.

Starting from the quantisation rule (27.3), we can ascertain the general nature of the distribution of levels in the energy spectrum. Let ΔE be the distance between two neighbouring levels, i.e. levels whose quantum numbers n differ by unity. Since ΔE is small (for large n) compared with the energy itself of the levels, we can write, from (27.3),

$$\Delta E \oint (\partial p/\partial E) \, \mathrm{d}x = 2\pi\hbar.$$

[†] A more exact analysis, using the exact (not quasi-classical) solutions of Schrödinger's equation near the turning points, gives $n+\frac{1}{2}$ instead of the integer n in (27.2) and (27.3). It is also found that the number of zeros of the wave function at finite distances in the entire region of the motion is precisely n.

But for classical motion $\partial E/\partial p = v$, the velocity of the particle, so that

$$\oint (\partial p/\partial E)\, \mathrm{d}x = \oint \mathrm{d}x/v = T.$$

Hence we have

$$\Delta E = 2\pi\hbar/T = \hbar\omega. \tag{27.4}$$

Thus the distance between two neighbouring levels is $\hbar\omega$. The frequencies ω may be regarded as approximately the same for several adjacent levels (the difference in whose numbers n is small compared with n itself). Hence we reach the conclusion that, in any small range of a quasi-classical part of the spectrum, the levels are equidistant, at intervals of $\hbar\omega$. This result could have been foreseen, since, in the quasi-classical case, the frequencies corresponding to transitions between different energy levels must be integral multiples of the classical frequency ω.

It is of interest to investigate what the matrix elements of any physical quantity f become in the limit of classical mechanics. To do this, we start from the fact that the mean value $\bar{f}$ in any quantum state must become, in the limit, simply the classical value of the quantity, provided that the state itself gives, in the limit, a motion of the particle in a definite path. A wave packet (see §6) corresponds to such a state; it is obtained by superposition of a number of stationary states with nearly the same energy. The wave function of such a state is of the form

$$\Psi = \sum_n a_n \Psi_n,$$

where the coefficients a_n are noticeably different from zero only in some range Δn of values of the quantum number n such that $1 \ll \Delta n \ll n$; the numbers n are supposed large, because the stationary states are quasi-classical. The mean value of f is, by definition,

$$\bar{f} = \int \Psi^* \hat{f} \Psi\, \mathrm{d}x = \sum_n \sum_m a_m^* a_n f_{mn}\, e^{i\omega_{mn}t},$$

or, replacing the summation over n and m by a summation over n and the difference $m-n = s$,

$$\bar{f} = \sum_n \sum_s a^*_{n+s} a_n f_{n+s,\, n}\, e^{i\omega st},$$

where we have put $\omega_{mn} = s\omega$ in accordance with (27.4).

The matrix elements f_{nm} calculated by means of the quasi-classical wave functions decrease rapidly in magnitude as the difference $m-n$ increases, though at the same time they vary only slowly with n itself ($m-n$ being fixed). Hence we can write approximately

$$\bar{f} = \sum_n \sum_s a^*_n a_n f_s\, e^{i\omega st} = \sum_n |a_n|^2 \sum_s f_s\, e^{i\omega st},$$

where we have introduced the notation $f_s = f_{\bar{n}+s,\, \bar{n}}$, $\bar{n}$ being some mean value of the quantum number in the range Δn. But $\sum |a_n|^2 = 1$; hence

$$\bar{f} = \sum_s f_s\, e^{i\omega st}. \tag{27.5}$$

The sum obtained is in the form of an ordinary Fourier series. Since $\bar{f}$ must, in the limit, coincide with the classical quantity $f(t)$, we arrive at the result that the matrix elements f_{mn} in the limit become the components f_{m-n} in the expansion of the classical function $f(t)$ as a Fourier series.

The relation (27.3) can also be interpreted in another manner. The integral $\oint p\, \mathrm{d}x$ is the area enclosed by the closed classical phase trajectory of the particle (i.e. the curve in the px-plane, which is the phase space of the particle). Dividing this area into cells, each of area $2\pi\hbar$, we have n cells altogether; n, however, is the number of states with energies not exceeding the given value (corresponding to the phase trajectory considered). Thus we can say that, in the quasi-classical case, there corresponds to each quantum state a *cell in phase space* of area $2\pi\hbar$. In other words, the number of states belonging to the volume element $\Delta p \Delta x$ of phase space is

$$\Delta p\, \Delta x / 2\pi\hbar. \tag{27.6}$$

If we introduce, instead of the momentum, the *wave number* $k = p/\hbar$, this number can be written

$$\Delta k\, \Delta x / 2\pi.$$

It is, as we should expect, the same as the familiar expression for the number of proper vibrations of a wave field (see *Mechanics and Electrodynamics*, §78).

The important concept of cells in phase space is valid for any quasi-classical motion, and not only for the one-dimensional motion discussed here. This is clear from the relationship noted above with the number of proper vibrations of the wave field in a given volume of space. In the general case of a system with s degrees of freedom, there are

$$\Delta q_1 \dots \Delta q_s \, \Delta p_1 \dots \Delta p_s / (2\pi\hbar)^s \tag{27.7}$$

quantum states in a volume element in phase space. In particular, a free motion in a sufficiently large volume Ω is always quasi-classical.[†] The number of quantum states for such a motion with momentum components in specified ranges Δp_x, Δp_y, Δp_z is

$$\Omega \cdot \Delta p_x \, \Delta p_y \, \Delta p_z / (2\pi\hbar)^3. \tag{27.8}$$

The concept of a particle moving in a large but bounded region Ω is sometimes used in order to deal with a discrete instead of a continuous spectrum, thus simplifying the formulae; this procedure will be used in Part II. For motion in a bounded volume, the eigenvalues of the momentum components take a discrete series of values, the intervals between which are inversely proportional to the linear dimensions of the region, and tend to zero as these increase. The number density of these states is given by (27.8). The normalised wave functions (plane waves) of the stationary states of a discrete spectrum of this kind have the form

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} e^{i\mathbf{p}\cdot\mathbf{r}} \tag{27.9}$$

and are said to be normalised to one particle in the volume Ω.

[†] Where a "normalisation volume" is needed, it will always be denoted by Ω. This is a fictitious quantity, which never appears in the final physical results, and is introduced only to facilitate the discussion.

§28. The transmission coefficient

Let us consider the motion of a particle in one dimension in a field of the type shown in Fig. 4: $U(x)$ increases monotonically from one constant limit ($U = 0$ as $x \to -\infty$) to another ($U = U_0$ as $x \to +\infty$). According to classical mechanics, a particle of energy $E < U_0$ moving in such a field from left to right, on reaching such a "potential wall", is reflected from it, and begins to move in the opposite direction; if, however, $E > U_0$, the particle continues to move in its original direction, though with diminished velocity. In quantum mechanics, a new phenomenon appears: even for $E > U_0$, the particle may be "reflected" from the potential wall. The probability of reflection must in principle be calculated as follows.

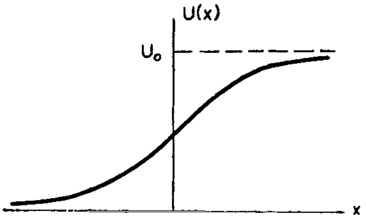

FIG. 4

Let the particle be moving from left to right. For large positive values of x, the wave function must describe a particle which has passed "above the wall" and is moving in the positive direction of x, i.e. it must have the asymptotic form

$$\text{for} \quad x \to \infty, \quad \psi \approx A\, e^{ik_2 x}, \quad \text{where} \quad k_2 = (1/\hbar)\sqrt{[2m(E-U_0)]} \tag{28.1}$$

and A is a constant. To find the solution of Schrödinger's equation which satisfies this boundary condition, we calculate the asymptotic expression for $x \to -\infty$; it is a linear combination of the two solutions of the equation of free motion, i.e. it has the form

$$\text{for} \quad x \to -\infty, \quad \psi \approx e^{ik_1 x} + Be^{-ik_1 x}, \quad \text{where} \quad k_1 = \sqrt{(2mE)}/\hbar. \tag{28.2}$$

The first term corresponds to a particle incident on the "wall" (we suppose ψ normalised so that the coefficient of this term is unity); the second term represents a particle reflected from the "wall". The current density in the incident wave is k_1, in the reflected wave $k_1 |B|^2$, and in the transmitted wave $k_2 |A|^2$. We define the *transmission coefficient D* as the ratio of the current density in the transmitted wave to that in the incident wave:

$$D = (k_2/k_1) |A|^2. \tag{28.3}$$

Similarly we can define the *reflection coefficient R* as the ratio of the density in the reflected wave to that in the incident wave. Evidently $R = 1-D$:

$$R = |B|^2 = 1-(k_2/k_1) |A|^2 \tag{28.4}$$

(this relation between A and B is automatically satisfied).

If the particle moves from left to right with energy $E < U_0$, then k_2 is purely imaginary, and the wave function decreases exponentially beyond the wall. The reflected current is equal to the incident one, i.e. we have total reflection of the particle from the potential wall.

A similar treatment can be used for passage through a *potential barrier*, i.e. a region of space in which the potential energy exceeds the total energy of the particle (Fig. 5 shows a one-dimensional barrier). It has already been mentioned in §22 that in quantum mechanics a particle reaching the barrier has a non-zero probability of passing "through" it. The permeability of the barrier to incident particles may be represented by a transmission coefficient, again defined as the ratio of the transmitted current density to the incident current density.

This coefficient can be estimated in a general form for a one-dimensional barrier satisfying the quasi-classical condition. According to this condition (see (26.12)), the "classical momentum" $p(x)$ of the particle, and therefore the potential energy $U(x)$ itself, must vary sufficiently slowly with x. This means that a quasi-classical potential barrier must have only a small slope; it must therefore be wide, and so the transmission coefficient is small in the quasi-classical case.

Let the particle be incident on the barrier from the left (from region I in Fig. 5). In the "classically inaccessible" region II, the wave

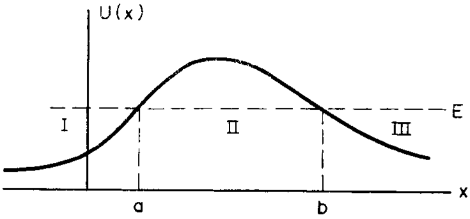

FIG. 5

function decreases exponentially from left to right:

$$\psi \sim \exp\left(-\frac{1}{\hbar}\int\limits_a^x |p|\,\mathrm{d}x\right), \qquad |p| = \sqrt{[2m(U-E)]}$$

(cf. (26.11)); the relatively slowly varying non-exponential factors are omitted here and henceforward. At the other side of the barrier ($x = b$), the wave function is thus attenuated in the ratio

$$\exp\left(-\frac{1}{\hbar}\int\limits_a^b |p|\,\mathrm{d}x\right)$$

in comparison with its value in the incident wave at $x = a$. The current density is proportional to the squared modulus of the wave function (again ignoring slowly varying factors). The ratio of the current transmitted through the barrier to the incident current is therefore

$$D \sim \exp\left(-\frac{2}{\hbar}\int\limits_a^b |p|\,\mathrm{d}x\right). \tag{28.5}$$

This estimate of the barrier transmission coefficient remains valid in the more realistic cases where the barrier is quasi-classical over much but not all of its extent. Such cases include those in which the potential-energy curve has a low slope only on one side, and on the other side is so steep that the quasi-classical approximation is invalid. The general condition for formula (28.5) to be valid is that the exponent must be large.

PROBLEMS

Problem 1. Determine the reflection coefficient of a particle from a rectangular potential wall (Fig. 6); the energy of the particle $E > U_0$.

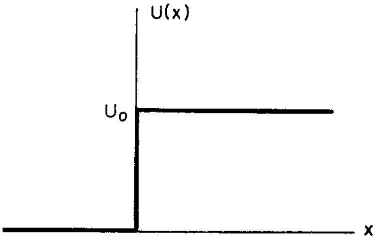

Fig. 6

Solution. Throughout the region $x > 0$, the wave function has the form (28.1), while in the region $x < 0$ its form is (28.2). The constants A and B are determined from the condition that ψ and $d\psi/dx$ are continuous at $x = 0$:

$$1+B = A, \quad k_1(1-B) = k_2 A,$$

whence

$$A = 2k_1/(k_1+k_2), \quad B = (k_1-k_2)/(k_1+k_2).$$

The reflection coefficient[†] is (28.4)

$$R = \left(\frac{k_1-k_2}{k_1+k_2}\right)^2 = \left(\frac{p_1-p_2}{p_1+p_2}\right)^2.$$

For $E = U_0$ ($k_2 = 0$), R becomes unity, while for $E \to \infty$ it tends to zero as $(U_0/4E)^2$.

Problem 2. Determine the transmission coefficient for a rectangular potential barrier (Fig. 7).

[†] In the limiting case of classical mechanics, the reflection coefficient must become zero. The expression obtained here, however, does not contain the quantum constant at all. This apparent contradiction is explained as follows. The classical limiting case is that in which the de Broglie wavelength of the particle $\lambda \sim \hbar/p$ is small in comparison with the characteristic dimensions of the problem, i.e. the distances over which the field $U(x)$ changes noticeably. In the schematic example considered, however, this distance is zero (at the point $x = 0$), so that the passage to the limit cannot be effected.

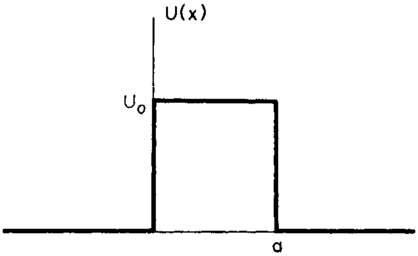

FIG. 7

SOLUTION. Let E be greater than U_0, and suppose that the incident particle is moving from left to right. Then we have for the wave function in the different regions expressions of the form

$$\begin{aligned} &\text{for} \quad x < 0, & \psi &= e^{ik_1x}+A\,e^{-ik_1x}, \\ &\text{for} \quad 0 < x < a, & \psi &= B\,e^{ik_2x}+B'\,e^{-ik_2x}, \\ &\text{for} \quad x > a, & \psi &= C\,e^{ik_1x} \end{aligned}$$

(on the side $x > a$ there can be only the transmitted wave, propagated in the positive direction of x). The constants A, B, B' and C are determined from the conditions of continuity of ψ and $d\psi/dx$ at the points $x = 0$ and a. The transmission coefficient is determined as $D = k_1|C|^2/k_1 = |C|^2$. On calculating this, we obtain

$$D = \frac{4k_1^2k_2^2}{(k_1^2-k_2^2)^2\sin^2 ak_2+4k_1^2k_2^2}.$$

For $E < U_0$, k_2 is a purely imaginary quantity; the corresponding expression for D is obtained by replacing k_2 by $i\varkappa_2$, where $\hbar\varkappa_2 = \sqrt{[2m(U_0-E)]}$:

$$D = \frac{4k_1^2\varkappa_2^2}{(k_1^2+\varkappa_2^2)^2\sinh^2 a\varkappa_2+4k_1^2\varkappa_2^2}.$$

PROBLEM 3. Determine from formula (28.5) the transmission coefficient for the potential barrier shown in Fig. 8: $U(x) = 0$ for $x < 0$, $U(x) = U_0-Fx$ for $x > 0$.

SOLUTION. A simple calculation gives the result

$$D \sim \exp\left[-\frac{4\sqrt{(2m)}}{3\hbar F}(U_0-E)^{3/2}\right].$$

PROBLEM 4. Determine the probability that a particle (with zero angular momentum) will emerge from a centrally symmetric potential well with $U(r) = -U_0$ for $r < r_0$, $U(r) = \alpha/r$ (Coulomb repulsion) for $r > r_0$ (Fig. 9).

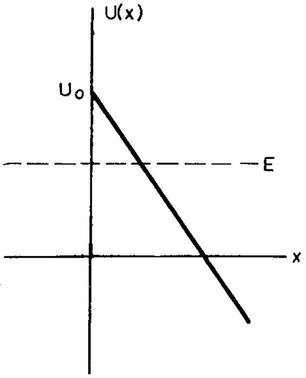

FIG. 8

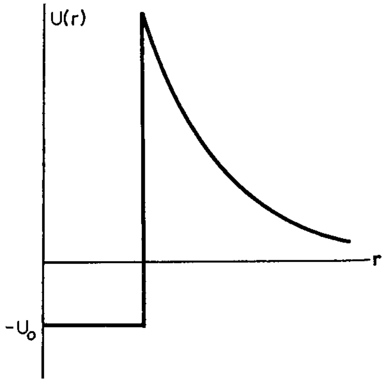

FIG. 9

SOLUTION. According to (28.5)†

$$w \sim \exp\left\{-\frac{2}{\hbar}\int_{r_0}^{\alpha/E}\sqrt{\left[2m\left(\frac{\alpha}{r}-E\right)\right]}\,dr\right\}.$$

Evaluating the integral, we finally obtain

$$w \sim \exp\left\{-\frac{2\alpha}{\hbar}\sqrt{\frac{2m}{E}}\left[\cos^{-1}\sqrt{\frac{Er_0}{\alpha}}-\sqrt{\left\{\frac{Er_0}{\alpha}\left(1-\frac{Er_0}{\alpha}\right)\right\}}\right]\right\}.$$

---

† Here we use the fact that a problem of motion of a particle with zero angular momentum in a central field reduces to a problem of one-dimensional motion with the same potential energy (see §30).

In the limiting case $r_0 \to 0$, this formula becomes

$$w \sim e^{-(\pi\alpha/\hbar)\sqrt{(2m/E)}} = e^{-2\pi\alpha/\hbar v}.$$

These formulae are applicable when the exponent is large, i.e. when $\alpha/\hbar v \gg 1$.

§29. Motion in a centrally symmetric field

The problem of the motion of two interacting particles can be reduced in quantum mechanics to that of one particle, as can be done in classical mechanics (see *Mechanics and Electrodynamics*, §11). The Hamiltonian of the two particles (of masses m_1, m_2) interacting in accordance with the law $U(r)$ (where r is the distance between the particles) is of the form

$$\hat{H} = -\frac{\hbar^2}{2m_1}\triangle_1 - \frac{\hbar^2}{2m_2}\triangle_2 + U(r), \tag{29.1}$$

where $\triangle_1$ and $\triangle_2$ are the Laplacian operators with respect to the coordinates of the particles. Instead of the radius vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ of the particles, we introduce new variables $\mathbf{R}$ and $\mathbf{r}$:

$$\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1, \quad \mathbf{R} = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/(m_1+m_2); \tag{29.2}$$

$\mathbf{r}$ is the vector of the distance between the particles, and $\mathbf{R}$ the radius vector of their centre of mass. A simple calculation gives

$$\hat{H} = -\frac{\hbar^2}{2(m_1+m_2)}\triangle_R - \frac{\hbar^2}{2m}\triangle + U(r), \tag{29.3}$$

where $\triangle_R$ and $\triangle$ are the Laplacian operators with respect to the components of the vectors $\mathbf{R}$ and $\mathbf{r}$ respectively, m_1+m_2 is the total mass of the system, and $m = m_1m_2/(m_1+m_2)$ is the reduced mass. Thus the Hamiltonian falls into the sum of two independent parts. Hence we can look for $\psi(\mathbf{r}_1, \mathbf{r}_2)$ in the form of a product $\phi(\mathbf{R})\psi(\mathbf{r})$, where the function $\phi(\mathbf{R})$ describes the motion of the centre of mass (as a free particle of mass m_1+m_2), and $\psi(\mathbf{r})$ describes the relative motion of the particles (as a particle of mass m moving in the central field $U(r)$).

Schrödinger's equation for the motion of a particle in a central field is

$$\triangle\psi+(2m/\hbar^2)[E-U(r)]\psi = 0. \tag{29.4}$$

Using the familiar expression for the Laplacian operator in spherical polar coordinates, we can write this equation in the form

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}\right]$$
$$+\frac{2m}{\hbar^2}[E-U(r)]\psi = 0. \tag{29.5}$$

If we introduce here the operator $\hat{\mathbf{l}}^2$ (14.15) of the squared angular momentum we obtain

$$\frac{\hbar^2}{2m}\left[-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{\hat{\mathbf{l}}^2}{r^2}\psi\right]+U(r)\psi = E\psi. \tag{29.6}$$

The angular momentum is conserved during motion in a central field. We shall consider stationary states in which the angular momentum and its component have definite values l and m. The specification of these values determines the angular dependence of the wave functions. Accordingly, we seek solutions of equation (29.6) in the form

$$\psi = R(r)Y_{lm}(\theta, \phi). \tag{29.7}$$

The eigenfunction of the angular momentum satisfies the equation $\hat{\mathbf{l}}^2Y_{lm} = l(l+1)Y_{lm}$, and we therefore obtain for the *radial function* $R(r)$ the equation

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R}{\mathrm{d}r}\right)-\frac{l(l+1)}{r^2}R+\frac{2m}{\hbar^2}[E-U(r)]R = 0. \tag{29.8}$$

We note that this equation does not contain the value of $l_z = m$ at all, in accordance with the $(2l+1)$-fold degeneracy of the levels with respect to the direction of the angular momentum, with which we are already familiar.

Let us investigate the radial part of the wave functions. By the substitution

$$R(r) = \chi(r)/r \tag{29.9}$$

equation (29.8) is brought to the form

$$\frac{d^2\chi}{dr^2}+\left[\frac{2m}{\hbar^2}(E-U)-\frac{l(l+1)}{r^2}\right]\chi = 0. \tag{29.10}$$

We shall assume that the potential energy $U(r)$, if it becomes infinite as $r \to 0$, does so less rapidly than $1/r^2$, i.e.

$$r^2U(r) \to 0 \quad \text{as} \quad r \to 0. \tag{29.11}$$

This excludes the possibility of the particle's "falling" to the centre (in a field for which $U \to -\infty$ as $r \to 0$), already mentioned in the footnote to §22. Then the wave function, and therefore the probability density $|\psi|^2$, remain finite in all space, including the point $r = 0$. Hence it follows that $\chi = rR$ must vanish for $r = 0$:

$$\chi(0) = 0. \tag{29.12}$$

Equation (29.10) is formally identical with Schrödinger's equation for one-dimensional motion in a field of potential energy

$$U_l(r) = U(r)+\frac{\hbar^2}{2m}\,\frac{l(l+1)}{r^2}, \tag{29.13}$$

in which the second term may be called the *centrifugal energy*. Thus the problem of motion in a central field reduces to that of one-dimensional motion in a region bounded on one side (the boundary condition for $r = 0$). The normalisation condition for the functions χ is also "one-dimensional":

$$\int_0^\infty |R|^2 r^2\, dr = \int_0^\infty |\chi|^2\, dr = 1. \tag{29.14}$$

The solution of equation (29.10) with the boundary condition (29.12) is fully determined by specifying the (permissible) value of E.

Hence, for motion in a central field, the state is completely determined by the values of E, l and m. In other words, the energy, the angular momentum and the component of the angular momentum together form a complete set of physical quantities for such a motion.

The reduction of the problem of motion in a central field to a one-dimensional problem enables us to apply the oscillation theorem (see §22). We arrange the eigenvalues of the energy (discrete spectrum) for a given l in order of increasing magnitude, and give them numbers n_r, the lowest level being given the number $n_r = 0$. Then n_r determines the number of nodes of the radial part of the wave function for finite values of r (excluding the point $r = 0$). The number n_r is called the *radial quantum number*. The number l for motion in a central field is sometimes called the *azimuthal quantum number*, and m the *magnetic quantum number*.

There is an accepted notation for states with various values of the angular momentum l of the particle: they are denoted by Latin letters, as follows:

$$\begin{array}{ccccccccccc} l = & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \ldots \\ & s & p & d & f & g & h & i & k & \ldots \end{array} \qquad (29.15)$$

Let us determine the form of the radial function near the origin. For small r, we seek $R(r)$ in the form $R = \text{constant} \times r^s$. Substituting this in the equation

$$\mathrm{d}(r^2\, \mathrm{d}R/\mathrm{d}r)/\mathrm{d}r - l(l+1)R = 0,$$

which is obtained from (29.8) by multiplying by r^2 and taking the limit as $r \to 0$, using (29.11), we find

$$s(s+1) = l(l+1).$$

Hence

$$s = l \quad \text{or} \quad s = -(l+1).$$

The solution with $s = -(l+1)$ does not satisfy the necessary conditions; it becomes infinite for $r = 0$. Thus the solution with $s = l$ remains, i.e. near the origin the wave functions of states with a given l are proportional to r^l:

$$R_l \approx \text{constant} \times r^l. \qquad (29.16)$$

The probability of a particle's being at a distance between r and $r+\mathrm{d}r$ from the centre is determined by the value of $r^2|R|^2$ and is thus proportional to $r^{2(l+1)}$. We see that it becomes zero at the origin the more rapidly, the greater the value of l.

§30. Spherical waves

The plane wave (20.9) describes a stationary state of a free particle in which the particle has a definite momentum **p** (and energy $E = p^2/2m$). Let us now consider stationary states (*spherical waves*) in which the particle has a definite value, not only of the energy, but also of the absolute value and component of the angular momentum. Instead of the energy, it is convenient to introduce the *wave number*

$$k = \sqrt{(2mE)}/\hbar. \tag{30.1}$$

The wave function of a state with angular momentum l and projection thereof m has the form

$$\psi_{klm} = R_{kl}(r)Y_{lm}(\theta, \phi), \tag{30.2}$$

where the radial function is determined by the equation

$$R''_{kl}+\frac{2}{r}R'_{kl}+\left[k^2-\frac{l(l+1)}{r^2}\right]R_{kl} = 0 \tag{30.3}$$

(equation (29.8) with $U(r) \equiv 0$). The wave functions ψ_{klm} for a spectrum continuous as regards k satisfy the conditions of normalisation and orthogonality:

$$\int \psi^*_{k'l'm'}\psi_{klm}\,\mathrm{d}V = \delta_{ll'}\delta_{mm'}\delta(k'-k).$$

The orthogonality for different l, l' and m, m' is ensured by the angular functions. The radial functions must be normalised by the condition

$$\int_0^\infty r^2 R_{k'l}R_{kl}\,\mathrm{d}r = \delta(k'-k). \tag{30.4}$$

For $l = 0$, equation (30.3) can be written

$$\frac{\mathrm{d}^2(rR_{k0})}{\mathrm{d}r^2}+k^2rR_{k0} = 0; \tag{30.5}$$

its solution finite for $r = 0$ and normalised by the condition (30.4) is

$$R_{k0} = \sqrt{\frac{2}{\pi}}\frac{\sin kr}{r}. \tag{30.6}$$

To check the normalisation, we can write

$$\int\limits_0^\infty r^2R_{k'0}R_{k0}\,\mathrm{d}r = \frac{2}{\pi}\int\limits_0^\infty \sin \text{[illegible]} \sin kr\,\mathrm{d}r$$

$$= \frac{1}{\pi}\int\limits_0^\infty \cos(k'-k)r\,\mathrm{d}r+\frac{1}{\pi}\int\limits_0^\infty \cos(k+k')r\,\mathrm{d}r. \tag{30.7}$$

According to the formula

$$\int\limits_0^\infty \cos\alpha x\,\mathrm{d}x = \pi\delta(\alpha), \tag{30.8}$$

the first integral in (30.7) gives the required δ-function; the second integral is zero, since $k+k' \neq 0$.[†]

When $l \neq 0$, the functions R_{kl} are more complicated, but at large distances r they can differ from (30.6) only in the phase of the trigonometric factor, since when $r \to \infty$ the term $l(l+1)/r^2$ can be omitted from (30.3), which then becomes the same as the equation with $l = 0$ (but, since the resulting equation relates only to the region of large r, it is no longer possible to choose one of the two independent solutions from the condition of finiteness for $r = 0$). The change of phase relative to the case $l = 0$ is in fact found to be $\frac{1}{2}\pi$, and the

---

[†] Formula (30.8) can be derived from (12.9) by taking the real part of each side and replacing the integral from $-\infty$ to ∞ by twice the integral from 0 to ∞.

asymptotic form of a spherical wave at large distances is therefore[†]

$$R_{kl} \approx \sqrt{\frac{2}{\pi}}\,\frac{\sin\left(kr - \frac{1}{2}l\pi\right)}{r}. \tag{30.9}$$

A similar asymptotic expression for the radial part of the wave function is valid not only in free motion of a particle but also for motion (with positive energy) in any field decreasing sufficiently rapidly as $r \to \infty$.[‡] At large distances, both the field and the centrifugal energy may be neglected in Schrödinger's equation, and we again have an equation of the form (30.5) for R_{kl}. The general solution of this equation is

$$R_{kl} \approx \sqrt{\frac{2}{\pi}}\,\frac{1}{r}\sin\left(kr - \tfrac{1}{2}l\pi + \delta_l\right), \tag{30.10}$$

where δ_l is a constant *phase shift*; the term $\frac{1}{2}l\pi$ in the argument of the sine is added so that $\delta_l = 0$ when the field is absent. The constant phase shift δ_l is determined by the boundary condition (R_{kl} is finite when $r = 0$) governing the solution of the exact Schrödinger's equation, and cannot be calculated in a general form. The phase shifts δ_l are, of course, functions of both l and k, and are an important property of the eigenfunctions of the continuous spectrum.

Let us consider a free particle moving with a definite momentum $p = \hbar k$ in the direction of the positive z-axis. The wave function of such a particle is a plane wave:

$$\begin{aligned} \psi &= \text{constant} \times e^{ikz} \\ &= \text{constant} \times e^{ikr\cos\theta}. \end{aligned} \tag{30.11}$$

It can be expanded in terms of the wave functions ψ_{klm} of free motion with definite orbital angular momenta. Since the function (30.11) has

---

[†] The solution of equation (30.3) that is finite at $r = 0$ can be expressed in terms of a Bessel function of half-integral order:

$$R_{kl} = J_{l+\frac{1}{2}}(kr)/\sqrt{(kr)}.$$

The usual asymptotic expression for the Bessel functions leads to (30.9).

[‡] Namely, the field $U(r)$ must decrease more rapidly than $1/r$.

axial symmetry about the z-axis, its expansion can contain only functions independent of the angle ϕ, i.e. having $m = 0$. These functions are $\psi_{kl0} = \text{constant} \times P_l(\cos\theta) R_{kl}$, and the required expansion must therefore be

$$e^{ikz} = \sum_{l=0}^{\infty} a_l R_{kl}(r) P_l(\cos\theta), \tag{30.12}$$

where the a_l are constant coefficients.

To determine these coefficients, we multiply equation (30.12) by $P_l(\cos\theta)\sin\theta$ and integrate over θ. Since the polynomials P_l with different l are orthogonal, and their normalisation integral is

$$\int_0^\pi P_l^2(\cos\theta)\sin\theta\, d\theta = \frac{2}{2l+1}, \tag{30.13}$$

we find

$$\int_0^\pi e^{ikr\cos\theta} P_l(\cos\theta)\sin\theta\, d\theta = a_l \frac{2}{2l+1} R_{kl}(r). \tag{30.14}$$

The integral on the left is easily calculated in the region of large r, where all the higher-order terms in $1/r$ may be neglected. Integrating by parts with the variable $\mu = \cos\theta$, we have to this accuracy

$$\int_{-1}^{1} e^{ikr\mu} P_l(\mu)\, d\mu \approx \left[P_l(\mu)\frac{e^{ikr\mu}}{ikr} \right]_{-1}^{1} = \frac{e^{ikr} - (-1)^l e^{-ikr}}{ikr},$$

using also the well-known values $P_l(1) = 1$, $P_l(-1) = (-1)^l$. This expression can also be written as

$$\frac{2i^l}{kr}\sin\left(kr - \tfrac{1}{2}l\pi\right);$$

the equation (30.14) with R_{kl} from (30.9) then gives

$$a_l = \sqrt{\frac{\pi}{2}}\,\frac{i^l}{k}(2l+1). \tag{30.15}$$

With these coefficients, the expansion (30.12) at large distances r has the asymptotic form

$$e^{ikz} \approx \frac{1}{kr}\sum_{l=0}^{\infty} i^l(2l+1)\,P_l(\cos\theta)\sin\left(kr-\tfrac{1}{2}\,l\pi\right). \tag{30.16}$$

This will be used later in discussing the theory of particle scattering.

PROBLEMS

Problem 1. Determine the energy levels for the motion of a particle with angular momentum $l = 0$ in a centrally symmetric potential well:

$$U(r) = -U_0 \quad \text{for} \quad r < a, \qquad U(r) = 0 \quad \text{for} \quad r > a.$$

Solution. For $l = 0$ the wave functions depend only on r. Inside the well, Schrödinger's equation has the form

$$\frac{1}{r}\frac{d^2}{dr^2}(r\psi)+k^2\psi = 0, \qquad k = \frac{1}{\hbar}\sqrt{[2m(U_0-|E|)]}.$$

The solution finite for $r = 0$ is

$$\psi = A\,\frac{\sin kr}{r}.$$

For $r > a$, we have the equation

$$\frac{1}{r}\frac{d^2}{dr^2}(r\psi)-\varkappa^2\psi = 0, \qquad \varkappa = \frac{1}{\hbar}\sqrt{(2m|E|)}.$$

The solution vanishing at infinity is

$$\psi = A'\,e^{-\varkappa r}/r.$$

The condition of the continuity of the logarithmic derivative of $r\psi$ at $r = a$ gives

$$k\cot ka = -\varkappa = -\sqrt{[(2mU_0/\hbar^2)-k^2]}, \tag{1}$$

or

$$\sin ka = \pm\sqrt{(\hbar^2/2ma^2U_0)}ka. \tag{2}$$

This equation determines in implicit form the required energy levels (we must take those roots of the equation for which $\cot ka < 0$, as follows from (1)). The first of these levels (with $l = 0$) is at the same time the deepest of all energy levels whatsoever, i.e. it corresponds to the normal state of the particle.

If the depth U_0 of the potential well is small enough, there are no levels of negative energy, and the particle cannot "stay" in the well. This is easily seen from equation (2), by means of the following graphical construction. The roots of an equation of the form $\pm\sin x = \alpha x$ are given by the points of intersection of the line $y = \alpha x$ with the curves $y = \pm\sin x$, and we must take only those points of intersection for which $\cot x < 0$; the corresponding parts of the curve $y = \pm\sin x$ are shown in Fig. 10 by a continuous line. We see that, if α is sufficiently large (U_0

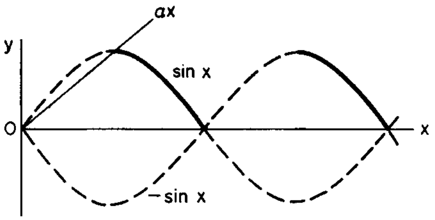

FIG. 10

small), there are no such points of intersection. The first such point appears when the line $y = \alpha x$ occupies the position shown, i.e. for $\alpha = 2/\pi$, and is at $x = \frac{1}{2}\pi$. Putting $\alpha = \hbar/\sqrt{(2ma^2U_0)}$, $x = ka$, we hence obtain for the minimum well depth to give a single negative level

$$U_{0,\,\mathrm{min}} = \pi^2\hbar^2/8ma^2. \tag{3}$$

This quantity is the greater, the smaller the well radius a. The position of the first level at the point where it first appears is determined from $ka = \frac{1}{2}\pi$ and is zero, as we should expect. As the well depth increases further, the normal level descends.

PROBLEM 2. Determine the energy levels of a three-dimensional oscillator (a particle in a field $U = \frac{1}{2}\mu\omega^2r^2$), and their degrees of degeneracy.

SOLUTION. Schrödinger's equation for a particle in a field $U = \frac{1}{2}m\omega^2(x^2+y^2+z^2)$ allows separation of the variables, leading to three equations like that of a linear oscillator. The energy levels are therefore

$$E_n = \hbar\omega(n_1+n_2+n_3+\tfrac{3}{2}) = \hbar\omega(n+\tfrac{3}{2}).$$

The degree of degeneracy of the nth level is equal to the number of ways in which n can be divided into the sum of three positive integral (or zero) numbers;† this is

$$\tfrac{1}{2}(n+1)\,(n+2).$$

§31. Motion in a Coulomb field

Let us consider the motion of an electron in a hydrogen atom or in a hydrogen-like ion, i.e. in the field of a nucleus whose charge is $+Ze$. If the nucleus is assumed to remain stationary, the problem reduces to that of the motion of a particle in an attractive Coulomb field:

$$U = -Ze^2/r. \tag{31.1}$$

---

† In other words, this is the number of ways in which n similar balls can be distributed among three urns.

From the general discussion in §22, it is evident from the start that the spectrum of positive eigenvalues of the energy E will be continuous, and that of negative energies discrete. We shall here be concerned with the latter, which corresponds to the bound states of the electron.

In problems connected with the Coulomb field it is convenient to use special units for the measurement of all quantities, known as *atomic units*. The units of measurement of mass, length and time are taken to be

$$m = 9.11\times10^{-28}\text{ g},\quad \hbar^2/me^2 = 0.529\times10^{-8}\text{ cm},$$
$$\hbar^3/me^4 = 2.42\times10^{-17}\text{ sec}$$

(where m is the mass of the electron); the atomic unit of length is called the *Bohr radius*. All the remaining units are expressed in terms of these; for example, the unit of energy[†] is

$$me^4/\hbar^2 = 4.36\times10^{-11}\text{ erg} = 27.21\text{ eV}.$$

The atomic unit of charge is the elementary charge $e = 4.80\times10^{-10}$ e.s.u. The formulae in atomic units can be obtained by putting $e = m = \hbar = 1$.

Equation (29.8) for the radial functions has the form

$$\frac{d^2R}{dr^2}+\frac{2}{r}\frac{dR}{dr}-\frac{l(l+1)}{r^2}R+\frac{2m}{\hbar^2}\left(E+\frac{Ze^2}{r}\right)R = 0 \tag{31.2}$$

or, in the new units,

$$\frac{d^2R}{dr^2}+\frac{2}{r}\frac{dR}{dr}-\frac{l(l+1)}{r^2}R+2\left(E+\frac{Z}{r}\right)R = 0. \tag{31.3}$$

Instead of the parameter E and the variable r, we introduce the new quantities

$$n = Z/\sqrt{(-2E)},\qquad \varrho = 2rZ/n. \tag{31.4}$$

For negative E, n is a real positive number. This substitution converts

---

[†] A quantity equal to one half of this unit is called a *rydberg*.

(31.3) to the form

$$R''+\frac{2}{\varrho}R'+\left[-\frac{1}{4}+\frac{n}{\varrho}-\frac{l(l+1)}{\varrho^2}\right]R = 0 \qquad (31.5)$$

(the primes denote differentiation with respect to ϱ).

For small ϱ, the solution which satisfies the necessary conditions of finiteness is proportional to ϱ^l (see (29.16)). To calculate the asymptotic behaviour of R for large ϱ, we omit from (31.5) the terms in $1/\varrho$ and $1/\varrho^2$ and obtain the equation

$$R'' = \tfrac{1}{4}R,$$

whence $R = e^{\pm\varrho/2}$. The solution in which we are interested, which vanishes at infinity, consequently behaves as $e^{-\varrho/2}$ for large ϱ.

It is therefore natural to make the substitution

$$R = \varrho^l e^{-\varrho/2}w(\varrho), \qquad (31.6)$$

when equation (31.5) becomes

$$\varrho w''+(2l+2-\varrho)\,w'+(n-l-1)\,w = 0. \qquad (31.7)$$

The solution of this equation must diverge at infinity not more rapidly than every finite power of ϱ, while for $\varrho = 0$ it must be finite.

Proceeding exactly as in §25, we seek the solution as a series

$$w = \sum_{s=0}^{\infty} a_s\varrho^s; \qquad (31.8)$$

substitution in (31.7) gives

$$\sum_{s=1}^{\infty} [a_s s(s-1)+(2l+2)\,a_s s]\varrho^{s-1}+\sum_{s=0}^{\infty} [-a_s s+a_s(n-l-1)]\,\varrho^s = 0$$

or, replacing s by $s+1$ in the first sum,

$$\sum_{s=0}^{\infty} [a_{s+1}(s+1)\,(s+2l+2)+a_s(n-l-1-s)]\,\varrho^s = 0.$$

Equating to zero the coefficient of each power of ϱ, we obtain the recurrence relation

$$a_{s+1} = -a_s \frac{n-l-1-s}{(s+1)(s+2l+2)}. \tag{31.9}$$

Hence the series (31.8) reduces to a polynomial (of degree $n-l-1$) if $n = l+1, l+2, \ldots$.

Thus the number n must be a positive integer, and for a given l we must have

$$n \geqslant l+1. \tag{31.10}$$

Recalling the definition (31.4) of the parameter n, we find

$$E = -Z^2/2n^2, \quad n = 1, 2, \ldots. \tag{31.11}$$

This solves the problem of determining the energy levels of the discrete spectrum in a Coulomb field. We see that there are an infinite number of levels between the normal level $E_1 = -\frac{1}{2}$ and zero. The distances between successive levels diminish as n increases; the levels become more crowded as we approach the value $E = 0$, where the discrete spectrum closes up into the continuous spectrum. In ordinary units, formula (31.11) is[†]

$$E = -Z^2 me^4/2\hbar^2 n^2. \tag{31.12}$$

The integer n is called the *principal quantum number*. The radial quantum number defined in §29 is

$$n_r = n-l-1.$$

For a given value of the principal quantum number, l can take the values

$$l = 0, 1, \ldots, n-1, \tag{31.13}$$

i.e. n different values in all. Only n appears in the expression (31.11) for the energy. Hence all states with different l but the same n have

---

[†] Formula (31.12) was first derived by N. Bohr in 1913, before the discovery of quantum mechanics. In quantum mechanics it was derived by W. Pauli in 1926 using the matrix method, and a few months later by E. Schrödinger using the wave equation.

the same energy. Thus each eigenvalue is degenerate, not only with respect to the magnetic quantum number m (as in any motion in a centrally symmetric field) but also with respect to the number l. This latter degeneracy (called *accidental* or *Coulomb*) is a specific property of the Coulomb field. To each value of l there correspond, as we know, $2l+1$ different values of m. Hence the degree of degeneracy of the nth energy level is

$$\sum_{l=0}^{n-1}(2l+1) = n^2. \tag{31.14}$$

We shall not give here the general expression for the electron wave functions, but only the ground-state wave function. When $n = 1$ and $l = 0$, the series (31.8) reduces to a constant, and the same is true of the angular function Y_{00}. The wave function is therefore

$$\psi = \frac{Z^{3/2}}{\sqrt{\pi}} e^{-Zr}. \tag{31.15}$$

It is normalised by the usual condition

$$\int |\psi|^2 \, \mathrm{d}V \equiv 4\pi \int_0^\infty r^2 |\psi|^2 \, \mathrm{d}r = 1.$$

The "dimensions" of the atom are represented by the distance r at which there is a considerable decrease of the electron density $|\psi|^2$. For a hydrogen atom ($Z = 1$), this distance is, in order of magnitude, just the atomic unit of length, as we see from (31.15). In ordinary units, this is the Bohr radius $a_B = \hbar^2/me^2$. The order of magnitude of the velocity of the electron in the atom is given by the uncertainty relation: $mv \sim \hbar/a_B$, whence $v \sim e^2/\hbar$.

PROBLEMS

PROBLEM 1. Determine the probability distribution of various values of the momentum in the ground state of the hydrogen atom ($Z = 1$).

SOLUTION. The wave function in the $\mathbf{p}$ representation is given by (31.15) a the integral (12.12). The integral is calculated by changing to spherical pola coordinates with the polar axis along **p**:

$$a(\mathbf{p}) = \frac{1}{(2\pi)^{3/2}}\int \psi(\mathbf{r})e^{-i\mathbf{p}\cdot\mathbf{r}}\,\mathrm{d}V$$

$$= \frac{1}{\pi\sqrt{2}}\int\limits_{0}^{\infty}\int\limits_{-1}^{1} e^{-r\;-ipr\cos\theta}\,\mathrm{d}\cos\theta . r^2\,\mathrm{d}r.$$

The result is

$$a(\mathbf{p}) = \frac{2\sqrt{2}}{\pi}\,\frac{1}{(1+p^2)^2},$$

and the probability density in **p**-space is $|a(\mathbf{p})|^2$.

PROBLEM 2. Determine the mean potential of the field created by the nucleus and the electron in the ground state of the hydrogen atom.

SOLUTION. The mean potential ϕ_e created by an "electron cloud" at an arbitrary point **r** is most simply found as the spherically symmetric solution of Poisson's equation with charge density $\varrho = -|\psi|^2$:

$$\frac{1}{r}\,\frac{\mathrm{d}^2}{\mathrm{d}r^2}(r\phi_e) = -4\pi\varrho = 4e^{-2r}.$$

Integrating this equation, and choosing the constants so that $\phi_e(0)$ is finite and $\phi_e(\infty) = 0$, and adding the potential of the field of the nucleus, we obtain

$$\phi = \frac{1}{r}+\phi_e(r) = \left(\frac{1}{r}+1\right)e^{-2r}.$$

For $r \ll 1$ we have $\phi \approx 1/r$ (the field of the nucleus), and for $r \gg 1$ the potential $\phi \approx e^{-2r}$ (the nucleus is screened by the electron).

CHAPTER 4

PERTURBATION THEORY

§32. Perturbations independent of time

The exact solution of Schrödinger's equation can be found only in a comparatively small number of the simplest cases. The majority of problems in quantum mechanics lead to equations which are too complex to be solved exactly. Often, however, quantities of different orders of magnitude appear in the conditions of the problem; among them there may be small quantities such that, when they are neglected, the problem is so much simplified that its exact solution becomes possible. In such cases, the first step in solving the physical problem concerned is to solve exactly the simplified problem, and the second step is to calculate approximately the errors due to the small terms that have been neglected in the simplified problem. There is a general method of calculating these errors; it is called *perturbation theory*.

Let us suppose that the Hamiltonian of a given physical system is of the form

$$\hat{H} = \hat{H}_0 + \hat{V},$$

where $\hat{V}$ is a small correction (or *perturbation*) to the *unperturbed* operator $\hat{H}_0$. In §§32, 33 we shall consider perturbations which do not depend explicitly on time (the same is assumed regarding $\hat{H}_0$ also). The conditions which are necessary for it to be permissible to regard the operator $\hat{V}$ as "small" compared with the operator $\hat{H}_0$ will be derived below.

The problem of perturbation theory for a discrete spectrum can be formulated as follows. It is assumed that the eigenfunctions $\psi_n^{(0)}$ and eigenvalues $E_n^{(0)}$ of the discrete spectrum of the unperturbed operator $\hat{H}_0$ are known, i.e. the exact solutions of the equation

$$\hat{H}_0\psi^{(0)} = E^{(0)}\psi^{(0)} \tag{32.1}$$

are known. It is desired to find approximate solutions of the equation

$$\hat{H}\psi = (\hat{H}_0+\hat{V})\psi = E\psi, \tag{32.2}$$

i.e. approximate expressions for the eigenfunctions ψ_n and eigenvalues E_n of the perturbed operator $\hat{H}$.

In this section we shall assume that no eigenvalue of the operator $\hat{H}_0$ is degenerate. Moreover, to simplify our results, we shall suppose that there is only a discrete spectrum of eigenvalues.

The calculations are conveniently performed in matrix form throughout. To do this, we expand the required function ψ in terms of the functions $\psi_n^{(0)}$:

$$\psi = \sum_m c_m\psi_m^{(0)}. \tag{32.3}$$

Substituting this expansion in (32.2) we obtain

$$\sum_m c_m(E_m^{(0)}+\hat{V})\psi_m^{(0)} = \sum_m c_m E\psi_m^{(0)};$$

multiplying both sides of this equation by $\psi_k^{(0)*}$ and integrating, we find

$$(E-E_k^{(0)})c_k = \sum_m V_{km}c_m. \tag{32.4}$$

Here we have introduced the matrix V_{km} of the perturbation operator, defined with respect to the unperturbed functions $\psi_m^{(0)}$:

$$V_{km} = \int \psi_k^{(0)*}\hat{V}\psi_m^{(0)}\,\mathrm{d}q. \tag{32.5}$$

We shall seek the values of the coefficients c_m and the energy E in the form of series

$$E = E^{(0)}+E^{(1)}+E^{(2)}+\dots, \qquad c_m = c_m^{(0)}+c_m^{(1)}+c_m^{(2)}+\dots,$$

where the quantities $E^{(1)}$ and $c_m^{(1)}$ are of the same order of smallness as the perturbation $\hat{V}$, the quantities $E^{(2)}$ and $c_m^{(2)}$ are of the second order of smallness, and so on.

Let us determine the corrections to the nth eigenvalue and eigenfunction, putting accordingly $c_n^{(0)} = 1$, $c_m^{(0)} = 0$ for $m \neq n$. To find the first approximation, we substitute in equation (32.4) $E = E_n^{(0)} + E_n^{(1)}$, $c_k = c_k^{(0)} + c_k^{(1)}$, and retain only terms of the first order. The equation with $k = n$ gives

$$E_n^{(1)} = V_{nn} = \int \psi_n^{(0)*} \hat{V} \psi_n^{(0)} \, dq. \tag{32.6}$$

Thus the first-order correction to the eigenvalue $E_n^{(0)}$ is equal to the mean value of the perturbation in the state $\psi_n^{(0)}$.

The equation (32.4) with $k \neq n$ gives

$$c_k^{(1)} = V_{kn}/(E_n^{(0)} - E_k^{(0)}) \quad \text{for} \quad k \neq n, \tag{32.7}$$

while $c_n^{(1)}$ remains arbitrary; it must be chosen so that the function $\psi_n = \psi_n^{(0)} + \psi_n^{(1)}$ is normalised up to and including terms of the first order. For this we can put $c_n^{(1)} = 0$. For the functions

$$\psi_n^{(1)} = \sum_m{}' \frac{V_{mn}}{E_n^{(0)} - E_m^{(0)}} \psi_m^{(0)} \tag{32.8}$$

(the prime means that the term with $m = n$ is omitted from the sum) are orthogonal to $\psi_n^{(0)}$, and hence the integral of $|\psi_n^{(0)} + \psi_n^{(1)}|^2$ differs from unity only by a quantity of the second order of smallness.

Formula (32.8) determines the correction to the wave functions in the first approximation. Incidentally, we see from this formula the condition for the applicability of the above method. This condition is that the inequality

$$|V_{mn}| \ll |E_n^{(0)} - E_m^{(0)}| \tag{32.9}$$

must hold, i.e. the matrix elements of the perturbation must be small compared with the corresponding differences between the unperturbed energy levels.

Next, let us determine the correction to the eigenvalue $E_n^{(0)}$ in the second approximation. To do this, we substitute in (32.4) $E = E_n^{(0)} +$

$E_n^{(1)}+E_n^{(2)}$, $c_k = c_k^{(0)}+c_k^{(1)}+c_k^{(2)}$, and examine the terms of the second order of smallness. The equation with $k = n$ gives

$$E_n^{(2)}c_n^{(0)} = \sum_m{}' V_{nm}c_m^{(1)},$$

whence

$$E_n^{(2)} = \sum_m{}' \frac{|V_{mn}|^2}{E_n^{(0)}-E_m^{(0)}} \tag{32.10}$$

(we have substituted $c_m^{(1)}$ from (32.7) and used the fact that, since the operator $\hat{V}$ is Hermitian, $V_{mn} = V_{nm}^*$).

We notice that the correction in the second approximation to the energy of the normal state is always negative; for, since $E_n^{(0)}$ then corresponds to the lowest value of the energy, all the terms in the sum (32.10) are negative.

The results obtained can be generalised at once to the case where the operator $\hat{H}_0$ has also a continuous spectrum (but the perturbation is applied, as before, to a state of the discrete spectrum). To do so, we need only add to the sums over the discrete spectrum the corresponding integrals over the continuous spectrum.

For states of the continuous spectrum, of course, the question of a change in the energy levels does not arise, and only the corrections to the eigenfunctions can be calculated.

In this connection, mention should be made of the case where the perturbation is represented by the potential energy of a particle in a weak external field, i.e. in a sufficiently shallow potential well. The unperturbed Schrödinger's equation is then simply the equation of free motion of the particle, and the energy levels are positive and form a continuous spectrum.

It has been shown at the end of §24 that there are no bound states (i.e. no negative energy levels) in such a well. When the energy E is zero, the unperturbed wave function $\psi^{(0)}$ of free motion reduces to a constant. Since the correction $\psi^{(1)} \ll \psi^{(0)}$, it is clear that the perturbed wave function $\psi = \psi^{(0)}+\psi^{(1)}$ of the motion in the well is nowhere zero. An eigenfunction without nodes belongs to the normal state

(§22). Thus $E = 0$ remains the least possible value of the energy of the particle.

The condition for perturbation theory to be applicable in this case requires that the well depth $|U|$ is small in comparison with the mean kinetic energy which the particle would have if enclosed within the well. According to the uncertainty relation, the momentum of such a particle would be $p \sim \hbar/a$, where a is the linear dimension of the particle; this leads to the condition $|U| \ll \hbar^2/ma^2$ stated in §24.[†]

PROBLEM

Determine the energy levels of an anharmonic linear oscillator whose Hamiltonian is

$$\hat{H} = \tfrac{1}{2}\hat{p}^2/m + \tfrac{1}{2}x^2\omega^2 m + \alpha x^3 + \beta x^4.$$

SOLUTION. The matrix elements of x^3 and x^4 can be obtained directly according to the rule of matrix multiplication, using the expression (25.4) for the matrix elements of x. We find for the matrix elements of x^3 that are not zero

$$(x^3)_{n-3,n} = (x^3)_{n,n-3} = (\hbar/2m\omega)^{3/2}\sqrt{[n(n-1)(n-2)]},$$
$$(x^3)_{n-1,n} = (x^3)_{n,n-1} = 3(\hbar/2m\omega)^{3/2}\, n^{3/2}.$$

The diagonal elements in this matrix vanish, so that the correction in the first approximation due to the term αx^3 in the Hamiltonian (regarded as a perturbation of the harmonic oscillator) is zero. The correction in the second approximation due to this term is of the same order as that in the first approximation due to the term βx^4. The diagonal matrix elements of x^4 are

$$(x^4)_{n,n} = \tfrac{3}{4}(\hbar/m\omega)^2(2n^2+2n+1).$$

Using the general formulae (32.6) and (32.10), we find the following approximate expression for the energy levels of the anharmonic oscillator:

$$E_n = \hbar\omega\left(n+\frac{1}{2}\right) - \frac{15}{4}\frac{\alpha^2}{\hbar\omega}\left(\frac{\hbar}{m\omega}\right)^3\left(n^2+n+\frac{11}{30}\right) + \frac{3}{2}\beta\left(\frac{\hbar}{m\omega}\right)^2\left(n^2+n+\frac{1}{2}\right).$$

[†] A one-dimensional or two-dimensional well, in which the field depends on only one or two coordinates, has infinite dimensions in two directions or one direction respectively, so that this condition cannot be satisfied. This is the reason for the inapplicability of perturbation theory to low-energy motion in such wells, and therefore for the invalidity of the deduction that there are no bound states.

§33. The secular equation

Let us now turn to the case where the unperturbed operator $\hat{H}_0$ has degenerate eigenvalues. We denote by $\psi_n^{(0)}$, $\psi_{n'}^{(0)}, \ldots$ the eigenfunctions belonging to the same eigenvalue $E_n^{(0)}$ of the energy. The choice of these functions is, as we know, not unique; instead of them we can choose any s (where s is the degree of degeneracy of the level $E_n^{(0)}$) independent linear combinations of these functions. The choice ceases to be arbitrary, however, if we subject the wave functions to the requirement that the change in them under the action of the small applied perturbation should be small.

At present we shall understand by $\psi_n^{(0)}$, $\psi_{n'}^{(0)}, \ldots$ some arbitrarily selected unperturbed eigenfunctions. The correct functions in the zeroth approximation are linear combinations of the form $c_n^{(0)}\psi_n^{(0)}+c_{n'}^{(0)}\psi_{n'}^{(0)}+\ldots$. The coefficients in these combinations are determined, together with the corrections in the first approximation to the eigenvalues, as follows.

We write out equations (32.4) with $k = n, n', \ldots$, and substitute in them, in the first approximation, $E = E_n^{(0)}+E^{(1)}$; for the quantities c_k it suffices to take the zero-order values $c_n = c_n^{(0)}$, $c_{n'} = c_{n'}^{(0)}, \ldots$; $c_m = 0$ for $m \neq n, n', \ldots$. We then obtain

$$E^{(1)}c_n^{(0)} = \sum_{n'} V_{nn'}c_{n'}^{(0)}$$

or

$$\sum_{n'} (V_{nn'}-E^{(1)}\delta_{nn'})c_{n'}^{(0)} = 0, \tag{33.1}$$

where n, n' take all values denumerating states belonging to the given unperturbed eigenvalue $E_n^{(0)}$. This system of homogeneous linear equations for the quantities $c_n^{(0)}$ has solutions which are not all zero if the determinant of the coefficients of the unknowns vanishes. Thus we obtain the equation

$$|V_{nn'}-E^{(1)}\delta_{nn'}| = 0. \tag{33.2}$$

This equation is of the sth degree in $E^{(1)}$ and has, in general, s different real roots. These roots are the required corrections to the eigenvalues in the first approximation. Equation (33.2) is called the *secular equation.*[†]

Substituting in turn the roots of equation (33.2) in the system (33.1) and solving, we find the coefficients $c_n^{(0)}$ and so determine the eigenfunctions in the zeroth approximation.

As a result of the perturbation, an originally degenerate energy level ceases in general to be degenerate (the roots of equation (33.2) are in general distinct); the perturbation *removes* the degeneracy, as we say. The removal of the degeneracy may be either total or partial (in the latter case, after the perturbation has been applied, there remains a degeneracy of degree less than the original one).

PROBLEMS

PROBLEM 1. Determine the corrections to the eigenvalue in the first approximation and the correct functions in the zeroth approximation, for a doubly degenerate level.

SOLUTION. Equation (33.2) here has the form

$$\begin{vmatrix} V_{11}-E^{(1)} & V_{12} \\ V_{21} & V_{22}-E^{(1)} \end{vmatrix} = 0$$

(the suffixes 1 and 2 correspond to two arbitrarily chosen unperturbed eigenfunctions $\psi_1^{(0)}$ and $\psi_2^{(0)}$ of the doubly degenerate level in question). Solving, we find

$$E^{(1)} = \tfrac{1}{2}[(V_{11}+V_{22})\pm\hbar\omega^{(1)}], \qquad \hbar\omega^{(1)} = \sqrt{[(V_{11}-V_{22})^2+4\,|V_{12}|^2]}, \tag{1}$$

with the notation $\hbar\omega^{(1)}$ for the difference between the two values of $E^{(1)}$. Solving also equations (33.1) with these values of $E^{(1)}$, we obtain for the coefficients in the correct normalised functions in the zeroth approximation $\psi^{(0)} = c_1^{(0)}\psi_1^{(0)}+c_2^{(0)}\psi_2^{(0)}$ the values

$$\begin{aligned} c_1^{(0)} &= \left\{\frac{V_{12}}{2\,|V_{12}|}\left[1\pm\frac{V_{11}-V_{22}}{\hbar\omega^{(1)}}\right]\right\}^{1/2}, \\ c_2^{(0)} &= \pm\left\{\frac{V_{21}}{2\,|V_{12}|}\left[1\mp\frac{V_{11}-V_{22}}{\hbar\omega^{(1)}}\right]\right\}^{1/2}. \end{aligned} \tag{2}$$

PROBLEM 2. At the initial instant $t = 0$, a system is in a state $\psi_1^{(0)}$ which belongs to a doubly degenerate level. Determine the probability that, at a subsequent instant t, the system will be in the state $\psi_2^{(0)}$ with the same energy; the transition occurs under the action of a constant perturbation.

[†] The name is taken from celestial mechanics.

Solution. We form the correct functions in the zeroth approximation,

$$\psi = c_1\psi_1 + c_2\psi_2, \qquad \psi' = c_1'\psi_1 + c_2'\psi_2,$$

where c_1, c_2; c_1', c_2' are two pairs of coefficients determined by formulae (2) of Problem 1 (for brevity, we omit the index $^{(0)}$ on all quantities).

Conversely,

$$\psi_1 = \frac{c_2'\psi - c_2\psi'}{c_1c_2' - c_1'c_2}.$$

The functions ψ and ψ' belong to states with perturbed energies $E+E^{(1)}$ and $E+E^{(1)\prime}$, where $E^{(1)}$ and $E^{(1)\prime}$ are the two values of the correction (1) in Problem 1. On introducing the time factors we pass to the time-dependent wave functions:

$$\Psi_1 = \frac{e^{-(i/\hbar)Et}}{c_1c_2' - c_1'c_2}\left[c_2'\psi e^{-(i/\hbar)E^{(1)}t} - c_2\psi'^{\,-(i/\hbar)E^{(1)\prime}t}\right]$$

(at time $t = 0$, $\Psi_1 = \psi_1$). Finally, again expressing ψ, ψ' in terms of ψ_1, ψ_2, we obtain Ψ_1 as a linear combination of ψ_1 and ψ_2, with coefficients depending on time. The squared modulus of the coefficient of ψ_2 determines the required transition probability w_{12}. Calculation using (1) and (2) gives

$$w_{12} = \frac{2\,|V_{12}|^2}{(\hbar\omega^{(1)})^2}\{1 - \cos\omega^{(1)\prime}t\}.$$

We see that the probability varies periodically with time, with frequency $\omega^{(1)}$.

§34. Perturbations depending on time

Let us now go on to study perturbations depending explicitly on time. We cannot speak in this case of corrections to the eigenvalues, since, when the Hamiltonian is time-dependent (as will be the perturbed operator $\hat{H} = \hat{H}_0 + \hat{V}(t)$), the energy is not conserved, so that there are no stationary states. The problem here consists in approximately calculating the wave functions from those of the stationary states of the unperturbed system.

To do this, we shall apply a method analogous to the well-known method of varying the constants to solve linear differential equations.† Let $\Psi_k^{(0)}$ be the wave functions (including the time factor) of the

---

† The application of this method in quantum mechanics is due to P.A.M. Dirac (1926).

stationary states of the unperturbed system. Then an arbitrary solution of the unperturbed wave equation can be written in the form of a sum: $\Psi = \Sigma a_k \Psi_k^{(0)}$. We shall now seek the solution of the perturbed equation

$$i\hbar\, \partial\Psi/\partial t = (\hat{H}_0 + \hat{V})\,\Psi \tag{34.1}$$

in the form of a sum

$$\Psi = \sum_k a_k(t)\, \Psi_k^{(0)}, \tag{34.2}$$

where the expansion coefficients are functions of time. Substituting (34.2) in (34.1), and recalling that the functions $\Psi_k^{(0)}$ satisfy the equation

$$i\hbar\, \partial\Psi_k^{(0)}/\partial t = \hat{H}_0 \Psi_k^{(0)},$$

we obtain

$$i\hbar \sum_k \Psi_k^{(0)} \frac{\mathrm{d}a_k}{\mathrm{d}t} = \sum_k a_k \hat{V} \Psi_k^{(0)}.$$

Multiplying both sides of this equation on the left by $\Psi_m^{(0)*}$ and integrating, we have

$$i\hbar \frac{\mathrm{d}a_m}{\mathrm{d}t} = \sum_k V_{mk}(t)\, a_k, \tag{34.3}$$

where

$$V_{mk}(t) = \int \Psi_m^{(0)*} \hat{V} \Psi_k^{(0)}\, \mathrm{d}q = V_{mk}\, e^{i\omega_{mk}t}, \qquad \omega_{mk} = (E_m^{(0)} - E_k^{(0)})/\hbar$$

are the matrix elements of the perturbation, including the time factor (and it must be borne in mind that, when V depends explicitly on time, the quantities V_{mk} also are functions of time).

As the unperturbed wave function we take the wave function of the ith stationary state, for which the corresponding values of the coefficients in (34.2) are $a_i^{(0)} = 1$, $a_k^{(0)} = 0$ for $k \neq i$. To find the first approximation, we seek a_k in the form $a_k = a_k^{(0)} + a_k^{(1)}$, substituting $a_k = a_k^{(0)}$ on the right-hand side of equation (34.3), which already contains the small quantities V_{mk}. This gives

$$i\hbar\, \mathrm{d}a_k^{(1)}/\mathrm{d}t = V_{ki}(t). \tag{34.4}$$

In order to show the unperturbed function to which the correction is being calculated, we introduce a second suffix in the coefficients a_k, writing

$$\Psi_i = \sum_k a_{ki}(t)\, \Psi_k^{(0)}. \tag{34.5}$$

Accordingly, we write the result of integrating equation (34.4) in the form

$$a_{ki}^{(1)} = -(i/\hbar)\int V_{ki}(t)\, \mathrm{d}t = -(i/\hbar)\int V_{ki}\, e^{i\omega_{ki}t}\, \mathrm{d}t. \tag{34.6}$$

This gives the wave functions in the first approximation. The choice of limits in the integrals in (34.6) depends on the conditions of each specific problem. Let us suppose, for example, that the perturbation acts only during some finite interval of time (or that $V(t)$ diminishes sufficiently rapidly as $t \to \pm\infty$). Let the system be in the ith stationary state (of a discrete spectrum) before the perturbation begins to act (or in the limit as $t \to -\infty$). At any subsequent instant the state of the system will be determined by the function (34.5), where, in the first approximation,

$$\begin{aligned} a_{ki} &= a_{ki}^{(1)} = -\frac{i}{\hbar}\int_{-\infty}^{t} V_{ki}\, e^{i\omega_{ki}t}\, \mathrm{d}t \quad \text{for} \quad k \neq i, \\ a_{ii} &= 1 + a_{ii}^{(1)} = 1 - \frac{i}{\hbar}\int_{-\infty}^{t} V_{ii}\, \mathrm{d}t; \end{aligned} \tag{34.7}$$

the limits of integration are taken so that, as $t \to -\infty$, all the $a_{ki}^{(1)}$ tend to zero. After the perturbation has ceased to act (or in the limit $t \to +\infty$), the coefficients a_{ki} take constant values $a_{ki}(\infty)$, and the system is in the state with wave function

$$\Psi = \sum_k a_{ki}(\infty)\, \Psi_k^{(0)},$$

which again satisfies the unperturbed wave equation, but is different from the original function $\Psi_i^{(0)}$. According to the general rule, the squared modulus of the coefficient $a_{ki}(\infty)$ determines the probability for the system to have an energy $E_k^{(0)}$, i.e. to be in the kth stationary state.

Thus, under the action of the perturbation, the system may pass from its initial stationary state to any other. For uniformity, let the initial state be denoted by the subscript i, and the final state by the subscript f. The probability of the transition $i \rightarrow f$ is

$$W = \frac{1}{\hbar^2}\left|\int_{-\infty}^{\infty} V_{fi}\, e^{i\omega_{fi}t}\, \mathrm{d}t\right|^2. \tag{34.8}$$

If the perturbation $V(t)$ varies little during time intervals of the order of the period $1/\omega_{fi}$, the value of the integral in (34.8) will be very small, because of the presence in the integrand of the rapidly oscillating and variable-sign factor $e^{i\omega_{fi}t}$. In the limit when the applied perturbation varies arbitrarily slowly, the probability of any transition with change of energy (i.e. with a non-zero frequency ω_{fi}) tends to zero. Thus, when the perturbation changes sufficiently slowly (*adiabatically*), a system in any non-degenerate stationary state will remain in that state.

§35. Transitions in the continuous spectrum

One of the most important applications of perturbation theory is to calculate the probability of a transition between states of a continuous spectrum under the action of a constant (time-independent) perturbation. This includes various collision processes, where the system in the initial and final states is an assembly of colliding particles, and the perturbation is represented by the interaction between them. The phenomena to which the method described below applies also include processes where a system in a bound state disintegrates into freely moving parts. For definiteness, we shall first consider this latter case.†

Let ν denote the set of quantities which take a continuous sequence of values defining the states of the continuous spectrum, and $\mathrm{d}\nu$ the product of their differentials. The unperturbed wave functions of the

---

† Strictly speaking, the discrete-spectrum states of a system capable of disintegration are not stationary but quasi-stationary (see §38); this point is not important in the present discussion, but will be further treated in §102.

continuous spectrum will be supposed normalised by the δ-function on the ν scale (for example, the quantities ν may be the momentum components of the free particles; the wave functions must then be normalised by the δ-function of momentum). With this normalisation, the expansion of the wave function has, instead of (34.2), the form

$$\Psi = \sum_k a_k(t)\,\Psi_k^{(0)} + \int a_\nu(t)\Psi_\nu^{(0)}\,\mathrm{d}\nu, \tag{35.1}$$

where the sum is taken over the whole discrete spectrum, and the integration over the continuous spectrum; then $|a_\nu(t)|^2\,\mathrm{d}\nu$ is the probability of finding the system (at time t) in states in the interval from ν to $\nu+\mathrm{d}\nu$ (cf. §5).

At the instant $t = 0$, let the system be in the initial state, denoted by the subscript i, and let us find the probability of a transition of the system to the state f, in which the quantities ν have values in the interval $\mathrm{d}\nu_f$.

By an appropriate change of the subscripts in (34.6) and by integration (when V_{fi} is independent of time), we get

$$\begin{aligned} a_{fi} &= -\frac{i}{\hbar}\int_0^t V_{fi}e^{i\omega_{fi}t}\,\mathrm{d}t \\ &= V_{fi}\,\frac{1-e^{-i\omega_{fi}t}}{\hbar\omega_{fi}}. \end{aligned} \tag{35.2}$$

The lower limit of integration is chosen so that $a_{fi} = 0$ for $t = 0$, in accordance with the initial condition stated.

The squared modulus of (35.2) is

$$|a_{fi}|^2 = |V_{fi}|^2\,\frac{\sin^2 \frac{1}{2}\omega_{fi}t}{\hbar^2\omega_{fi}^2}. \tag{35.3}$$

It is easily seen that, when t is large, this function is proportional to t. This is proved by using the formula

$$\lim_{t\to\infty}\frac{\sin^2 \alpha t}{\pi t\alpha^2} = \delta(\alpha): \tag{35.4}$$

when $\alpha \neq 0$ the limit is zero, when $\alpha = 0$ we have $\sin^2 \alpha t/\alpha^2 t = t$ and the limit is infinite, and integration over α from $-\infty$ to ∞ with $\alpha t = \xi$ gives

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2 \alpha t}{t\alpha^2} \, \mathrm{d}\alpha = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2 \xi}{\xi^2} \, \mathrm{d}\xi = 1.$$

Thus the function on the left of (35.4) has all the properties of the δ-function. Accordingly, for large t we can write

$$|a_{fi}|^2 = \frac{1}{\hbar^2} |V_{fi}|^2 \pi t \delta \left(\tfrac{1}{2} \omega_{fi}\right)$$

or, since $\delta(\alpha x) = \delta(x)/\alpha$,

$$|a_{fi}|^2 = \frac{2\pi}{\hbar} |V_{fi}|^2 \, \delta(E_f - E_i) t. \tag{35.5}$$

The expression $|a_{fi}|^2 \, \mathrm{d}\nu_f$ is the probability of a transition from the original state to a state in the interval $\mathrm{d}\nu_f$. We see that, for large t, this expression is proportional to the time elapsed since $t = 0$. Without the factor t, it gives the transition probability $\mathrm{d}w$ per unit time (whose dimensions are $1/T$, in contrast to the dimensionless probability (34.7)):

$$\mathrm{d}w = \frac{2\pi}{\hbar} |V_{fi}|^2 \, \delta(E_f - E_i) \, \mathrm{d}\nu_f. \tag{35.6}$$

The probability is zero except for transitions to states with energy $E_f = E_i$, in accordance with the law of conservation of energy. This law is represented by the δ-function in (35.6), but the probability is, of course, not infinite (which would be meaningless) when $E_f = E_i$; in reality, the δ-function is eliminated by integration over a finite interval of states. For example, if the states of the continuous spectrum are not degenerate, $\mathrm{d}\nu_f$ must be taken as a single energy value. Then the integration of (35.6) over $\mathrm{d}\nu_f = \mathrm{d}E_f$ gives the value of the transition probability as

$$w = (2\pi/\hbar) \, |V_{fi}|^2. \tag{35.7}$$

Formula (35.6) is also applicable when the initial state of the system is also in the continuous spectrum, as occurs in the collision problem; an example will be given in §67. It must be noticed, however, that in such cases the quantity $\mathrm{d}w$ determined by formula (35.6) is not actually the transition probability; it does not even have the correct dimensions $(1/T)$. The expression (35.6) is proportional to the number of transitions per unit time, but its dimensions and sense depend on the way in which the initial wave functions of the continuous spectrum are normalised; for instance $\mathrm{d}w$ may be the collision cross-section, as will be seen in §67.

§36. Intermediate states

It may happen that the matrix element V_{fi} for the transition considered vanishes. Then formula (35.6) does not answer the question concerned, and in order to determine the transition probability we have to go to the next approximation of perturbation theory.

The correction $a_{fi}^{(1)}$ vanishes together with V_{fi}. For the second-order correction $a_{fi}^{(2)}$, equations (34.3) give

$$i\hbar\frac{\mathrm{d}a_{fi}^{(2)}}{\mathrm{d}t} = \sum_k V_{fk}e^{i\omega_{fk}t}a_{ki}^{(1)}, \tag{36.1}$$

where the summation is over the states for which the matrix elements for transitions $k \to f$ are non-zero. The first-order corrections $a_{ki}^{(1)}$ are given by the equations

$$i\hbar\frac{\mathrm{d}a_{ki}^{(1)}}{\mathrm{d}t} = V_{ki}e^{i\omega_{ki}t}$$

(cf. (34.4)); hence

$$a_{ki}^{(1)} = -(V_{ki}/\hbar\omega_{ki})(e^{i\omega_{ki}t}-1).$$

Substituting this in (36.1) and integrating, we obtain

$$a_{fi}^{(2)} = \frac{i}{\hbar^2}\sum_k \frac{V_{fk}V_{ki}}{\omega_{ki}}\int_0^t (e^{i\omega_{fi}t}-e^{i\omega_{fk}t})\,\mathrm{d}t.$$

In the integral we need retain only the first term, which will contain the small denominator ω_{fi}. Thus

$$a_{fi}^{(2)} = \left(\sum_k \frac{V_{fk} V_{ki}}{\hbar\omega_{ki}}\right) \frac{e^{i\omega_{fi}t}-1}{\hbar\omega_{fi}}.$$

This expression differs from (35.2) only in that the matrix element V_{fi} is replaced by the sum in parentheses. Accordingly, we obtain instead of (35.6)

$$\mathrm{d}w = \frac{2\pi}{\hbar} \left| \sum_k \frac{V_{fk} V_{ki}}{E_i - E_k} \right|^2 \delta(E_f - E_i)\, \mathrm{d}\nu_f. \tag{36.2}$$

In this connection the states k for which the matrix elements V_{fk} and V_{ki} are not zero are called *intermediate* states for the transition $i \rightarrow f$. We can say in an intuitive way that this transition takes place, as it were, in two stages: $i \rightarrow k$ and $k \rightarrow f$; such a description must not, of course, be understood literally.

§37. The uncertainty relation for energy

Let us consider a system composed of two weakly interacting parts. We suppose that it is known that at some instant these parts have definite values of the energy, which we denote by E and ε respectively. Let the energy be measured again after some time interval Δt; the values E', ε' obtained are in general different from E, ε. It is easy to determine the order of magnitude of the most probable value of the difference $E' + \varepsilon' - E - \varepsilon$ which is found as a result of the measurement.

According to formula (35.3), the probability of a transition of the system (after time t), under the action of a time-independent perturbation, from a state with energy E to one with energy E' is proportional to

$$\sin^2 [(E' - E)t/2\hbar]/(E' - E)^2.$$

Hence we see that the most probable value of the difference $E' - E$ is of the order of $\hbar/t$.

Applying this result to the case we are considering (the perturbation being the interaction between the parts of the system), we obtain the relation

$$|E+\varepsilon-E'-\varepsilon'|\,\Delta t \sim \hbar. \tag{37.1}$$

Thus the smaller the time interval Δt, the greater the energy change that is observed. It is important to notice that its order of magnitude $\hbar/\Delta t$ is independent of the amount of the perturbation. The energy change determined by the relation (37.1) will be observed, however weak the interaction between the two parts of the system. This result is peculiar to quantum theory and has a deep physical significance. It shows that, in quantum mechanics, the law of conservation of energy can be verified by means of two measurements only to an accuracy of the order of $\hbar/\Delta t$, where Δt is the time interval between the measurements.

The relation (37.1) is often called the *uncertainty relation for energy*. However, it must be emphasised that its significance is entirely different from that of the uncertainty relation $\Delta p\,\Delta x \sim \hbar$ for the coordinate and momentum. In the latter, Δp and Δx are the uncertainties in the values of the momentum and coordinate at the same instant; the relation shows that these two quantities can never have entirely definite values simultaneously. The energies E, ε, on the other hand, can be measured to any degree of accuracy at any instant. The quantity $(E+\varepsilon)-(E'+\varepsilon')$ in (37.1) is the difference between two exactly measured values of the energy $E+\varepsilon$ at two different instants, and not the uncertainty in the value of the energy at a given instant.

If we regard E as the energy of some system and ε as that of a "measuring apparatus", we can say that the energy of interaction between them can be taken into account only to within $\hbar/\Delta t$. Let us denote by ΔE, $\Delta\varepsilon$, . . . the errors in the measurements of the corresponding quantities. In the favourable case when ε, ε' are known exactly ($\Delta\varepsilon = \Delta\varepsilon' = 0$), we have

$$\Delta(E-E') \sim \hbar/\Delta t. \tag{37.2}$$

From this relation we can derive important consequences concerning the measurement of momentum. The process of measuring the momentum of a particle (for definiteness, we shall speak of an electron) consists in a collision of the electron with some other ("measuring") particle, whose momenta before and after the collision can be regarded as known exactly. The conservation laws for momentum and energy have to be applied to this process. The latter, however, can be applied, as we have seen, only to an accuracy of the order of $\hbar/\Delta t$, where Δt is the time between the beginning and end of the process in question.

To simplify the subsequent discussion, it is convenient to consider an imaginary idealised experiment in which the "measuring particle" is a perfectly reflecting plane mirror; only one momentum component is then of importance, namely that perpendicular to the plane of the mirror. To determine the momentum P of the particle, the laws of conservation of momentum and energy give the equations

$$p'+P'-p-P = 0, \tag{37.3}$$

$$|\varepsilon'+E'-\varepsilon-E| \sim \hbar/\Delta t, \tag{37.4}$$

where P, E are the momentum and energy of the particle, and p, ε those of the mirror; the unprimed and primed quantities refer to the instants before and after the collision respectively. The quantities p, p', ε, ε' relating to the "measuring particle" can be regarded as known exactly, i.e. the errors in them are zero. Then we have for the errors in the remaining quantities, from the above equations:

$$\Delta P = \Delta P', \quad \Delta E'-\Delta E \sim \hbar/\Delta t.$$

But $\Delta E = (\partial E/\partial P)\Delta P = v\Delta P$, where v is the velocity of the electron (before the collision), and similarly $\Delta E' = v'\,\Delta P' = v'\,\Delta P$. Hence we obtain

$$(v_x'-v_x)\Delta P_x \sim \hbar/\Delta t. \tag{37.5}$$

We have here added the suffix x to the velocity and momentum, in order to emphasise that this relation holds for each of their components separately.

This is the required relation. It shows that the measurement of the momentum of the electron (with a given degree of accuracy ΔP) necessarily involves a change in its velocity (i.e. in the momentum itself). This change is the greater, the shorter the duration of the measuring process. The change in velocity can be made arbitrarily small only as $\Delta t \to \infty$, but measurements of momentum occupying a long time can be significant only for a free particle. The non-repeatability of a measurement of momentum after short intervals of time, and the "two-faced" nature of measurement in quantum mechanics—the necessity of a distinction between the measured values of a quantity and the values resulting from the process of measurement—are here exhibited with particular clarity.[†]

§38. Quasi-stationary states

The conclusion reached at the beginning of §37 can also be derived from another standpoint by considering the decay of a system under the action of some perturbation. Let E_0 be some energy level of the system, calculated without any allowance for the possibility of its decay. We denote by τ the *lifetime* of this state of the system, i.e. the reciprocal of the probability w of decay per unit time:

$$\tau = 1/w. \tag{38.1}$$

Then we find by the same method that

$$|E_0 - E - \varepsilon| \sim \hbar/\tau,$$

where E, ε are the energies of the two parts into which the system decays. The sum $E+\varepsilon$, however, gives us an estimate of the energy of the system before it decays. Hence the above relation shows that the energy of a system which is free to decay can be determined only to within a quantity of the order of $\hbar/\tau$.

A system that is able to decay does not, strictly speaking, have a discrete energy spectrum. A particle leaving the system when it decays goes to infinity; in this sense, the motion of the system is infinite, and hence the energy spectrum is continuous.

[†] The relation (37.5) and the elucidation of the physical significance of the uncertainty relation for energy are due to N. Bohr (1928).

It may happen, however, that the decay probability of the system is very small. The simplest example of this kind is a particle surrounded by a sufficiently high and wide potential barrier. For such systems having a small decay probability, we can consider *quasi-stationary states*, in which the particles move for a considerable time "within the system" and leave it only much later. The energy spectrum of these states will be "quasi-discrete"; it consists of a series of broadened levels whose widths depend on their lifetimes. The width of a level can be quantitatively expressed by

$$\Gamma = \hbar/\tau = \hbar w. \tag{38.2}$$

The widths of the quasi-discrete levels are small compared with the distances between them.

In discussing the quasi-stationary states, we can use the following formal method. Until now we have always considered solutions of Schrödinger's equation with a boundary condition requiring the finiteness of the wave function at infinity. Instead of this, we shall now look for solutions which represent an outgoing spherical wave at infinity ($\psi \sim e^{ikr}/r$); this corresponds to a particle finally leaving the system when it decays. Since such a boundary condition is complex, we cannot assert that the eigenvalues of the energy must be real. On the contrary, by solving Schrödinger's equation we obtain a set of complex values, which we write in the form

$$E = E_0 - \tfrac{1}{2}i\Gamma, \tag{38.3}$$

where E_0 and Γ are two positive constants.

It is easy to see the physical significance of the complex energy values. The time factor in the wave function of a quasi-stationary state is of the form

$$e^{-(i/\hbar)Et} = e^{-(i/\hbar)E_0 t} e^{-\Gamma t/2\hbar}.$$

Hence all the probabilities given by the squared modulus of the wave function decrease with time as $e^{-\Gamma t/\hbar}$. In particular, the probability of finding the particle "within the system" decreases according to this law.

An extensive class of quasi-stationary states arises in nuclear reactions at not too high energies which pass through the stage of formation of a *compound nucleus*.[†] An intuitive physical picture of the processes occurring is that the particle (such as a neutron) incident on the nucleus interacts with the nucleons in the nucleus and "coalesces" with them, forming a compound system in which the energy contributed by the particle is distributed between many nucleons. The long lifetime of the quasi-stationary states in such a system (compared with the "periods" of the motion of the nucleons in the nucleus) is due to the fact that for the greater part of the time the energy is distributed between many particles, so that none of them has sufficient energy to overcome the attraction of the other particles and leave the nucleus. Sufficient energy to break up the compound nucleus is only comparatively rarely concentrated on one particle.

[†] The concept of the compound nucleus is due to N. Bohr (1936).

CHAPTER 5

SPIN

§39. Spin

Let us consider a composite particle, such as an atomic nucleus, at rest as a whole and in a definite internal state. In addition to an internal energy, it has also an angular momentum of definite magnitude L, due to the motion of the particles within the nucleus. For a given L the angular momentum can, as we know, have $2L+1$ different orientations in space.

It has been mentioned in §18 that an important feature of the concept of angular momentum in quantum mechanics is that this quantity determines the symmetry properties of the states of the system with respect to rotations in space. When the coordinates are rotated, the $2L+1$ wave functions ψ_{LM} corresponding to different values of the angular momentum component M are transformed into certain combinations of one another.

In this formulation, the origin of the angular momentum becomes unimportant, and we naturally arrive at the concept of an "intrinsic" angular momentum which must be ascribed to the particle regardless of whether it is "composite" or "elementary".

Thus, in quantum mechanics an elementary particle must be assigned a certain "intrinsic" angular momentum unconnected with its motion in space. This property of elementary particles is peculiar to quantum theory (it disappears in the limit $\hbar \to 0$), and therefore has in principle no classical interpretation.[†]

[†] In particular, it would be wholly meaningless to imagine the "intrinsic" angular momentum of an elementary particle as being the result of its rotation "about its own axis".

The intrinsic angular momentum of a particle is called its *spin*, as distinct from the angular momentum due to the motion of the particle in space, called the *orbital angular momentum*. The particle concerned may be either elementary, or composite but behaving in some respect as an elementary particle (e.g. an atomic nucleus). The spin of a particle (measured, like the orbital angular momentum, in units of $\hbar$) will be denoted by $\mathbf{s}$.[†]

For particles having spin, the description of the state by means of the wave function must determine the probability not only of its different positions in space but also of the possible orientations of the spin. Thus the wave function must depend not only on three continuous variables, the coordinates of the particle, but on a discrete *spin variable*, which gives the value of the projection of the spin on a selected direction in space (the z-axis) and takes a limited number of discrete values, which we shall denote by σ.

Let $\psi(x, y, z; \sigma)$ be such a wave function. It is essentially a set of several different functions of the coordinates, corresponding to different values of σ; these functions will be called the *spin components* of the wave function. The integral

$$\int |\psi(x, y, z; \sigma)|^2 \, \mathrm{d}V$$

determines the probability that the particle has a certain value of σ. The probability that the particle is in the volume element $\mathrm{d}V$ with any value of σ is

$$\sum_{\sigma} |\psi(x, y, z; \sigma)|^2 .$$

The quantum-mechanical spin operator, on being applied to the wave function, acts on the spin variable σ. In other words, it in some way transforms the components of the wave functions into linear combinations of one another. The form of this operator will be established later. However, it is clear *a priori* that the operators $\hat{s}_x$, $\hat{s}_y$, $\hat{s}_z$ of the three spin components satisfy the same commutation rules as the operators of the orbital angular momentum. The general defini-

---

[†] The physical idea that an electron has an intrinsic angular momentum was put forward by G. Uhlenbeck and S. Goudsmit in 1925. Spin was introduced into quantum mechanics in 1927 by W. Pauli.

tion of the angular momentum operators in quantum mechanics is based on their relation to the operators of infinitesimal rotations. In deriving the expressions and commutation rules for these operators, they were assumed to act on functions of the coordinates. In fact, however, these rules represent properties of rotations as such, no matter to what mathematical entity they are applied, and therefore have universal validity.

The commutation rules enable us to determine the possible values of the absolute magnitude and components of the spin. All the results derived in §15 (formulae (15.6)–(15.8)) were based only on the commutation rules, and hence are applicable here also; we need only replace **L** in these formulae by **s**. It follows from formulae (15.6) that the eigenvalues of the z-component of the spin form a sequence of numbers differing by unity. However, we cannot now assert that these values must be integral, as we could for the component l_z of the orbital angular momentum (the derivation given at the beginning of §15 is invalid here, since it was based on the specific expression for the operator $\hat{l}_z$, acting on functions of the coordinates).

Moreover, the sequence of eigenvalues s_z is limited above and below by values equal in absolute magnitude and opposite in sign, which we denote by $\pm s$. The difference $2s$ between the greatest and least values of s_z must be an integer or zero. Consequently s can take the values $0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$.

Thus the eigenvalues of the square of the spin are

$$\mathbf{s}^2 = s(s+1), \tag{39.1}$$

where s can be either an integer (including zero) or half an integer. For given s, the component s_z of the spin can take the values $\sigma = s, s-1, \ldots, -s$, i.e. $2s+1$ values in all. From what was said above, a particle whose spin is s must have a wave function with $2s+1$ components.[†]

[†] Since s is fixed for each kind of particle, the spin angular momentum $\hbar s$ becomes zero in the limit of classical mechanics ($\hbar \to 0$). This consideration does not apply to the orbital angular momentum, since l can take any value. The transition to classical mechanics is represented by $\hbar$ tending to zero and l simultaneously tending to infinity, in such a way that the product $\hbar l$ remains finite.

The majority of the elementary particles (electrons, protons, neutrons, μ-mesons) have a spin of $\frac{1}{2}$. There are also elementary particles with other spin values, for example the π-mesons and the K-mesons, whose spin is zero.

The total angular momentum of a particle, denoted by **j**, is composed of its orbital angular momentum **l** and its spin **s**. Their operators act on functions of different variables, and therefore, of course, commute. The eigenvalues of the total angular momentum

$$\mathbf{j} = \mathbf{l}+\mathbf{s} \tag{39.2}$$

are determined by the same "vector model" rule as the sum of the orbital angular momenta of two different particles (§17). That is, for given values of l and s, the total angular momentum can take the values $j = l+s,\ l+s-1,\ \ldots,\ |l-s|$. Thus, for an electron (spin $\frac{1}{2}$) with non-zero orbital angular momentum l, the total angular momentum can be $j = l\pm\frac{1}{2}$; for $l = 0$ the angular momentum j has, of course, only the one value $j = \frac{1}{2}$.

The operator of the total angular momentum **J** of a system of particles is equal to the sum of the operators of the angular momentum **j** of each particle, so that its values are again determined by the vector model rules. The angular momentum **J** can be put in the form $\mathbf{J} = \mathbf{L}+\mathbf{S}$, where **S** is the total spin and **L** the total orbital angular momentum of the particles.

The formulae (15.11) for the matrix elements of the angular momentum components are, like the commutation rules, valid universally (i.e. for any angular momentum). The angular momentum selection rules derived in §18 for the matrix elements of various physical quantities also remain valid (with appropriate change of notation).

§40. The spin operator

In §§40–42 we shall not be interested in the dependence of the wave function on the coordinates. For example, in speaking of the behaviour of the function $\psi(\sigma)$ when the system of coordinates is rotated, we can suppose that the particle is at the origin, so that its coordinates remain unchanged by such a rotation, and the results obtained will characterise the behaviour of the function ψ with regard to the spin variable σ.

The variable σ differs from the ordinary variables (the coordinates) by being discrete. The most general form of a linear operator acting on functions of a discrete variable σ is

$$\hat{f}\psi(\sigma) = \sum_{\sigma'} f_{\sigma\sigma'}\psi(\sigma'), \tag{40.1}$$

where the $f_{\sigma\sigma'}$ are constants.

It is easy to see that these quantities are the same as the matrix elements of the operator $\hat{f}$, determined by the usual rule (11.6) from the eigenfunctions of the operator $\hat{s}_z$. The integration over the coordinates in (11.6) is here replaced by summation over the discrete variable, so that the definition of the matrix element is

$$f_{\sigma_2\sigma_1} = \sum_{\sigma} \psi^*_{\sigma_2}(\sigma)\,[\hat{f}\psi_{\sigma_1}(\sigma)]. \tag{40.2}$$

Here $\psi_{\sigma_1}(\sigma)$ and $\psi_{\sigma_2}(\sigma)$ are the eigenfunctions of the operator $\hat{s}_z$ corresponding to the eigenvalues $s_z = \sigma_1$ and σ_2; each such function corresponds to a state in which the particle has a definite value of s_z, i.e. in which only one component of the wave function is non-zero:[†]

$$\psi_{\sigma_1}(\sigma) = \delta_{\sigma_1\sigma}, \quad \psi_{\sigma_2}(\sigma) = \delta_{\sigma_2\sigma}. \tag{40.3}$$

According to (40.1),

$$\hat{f}\psi_{\sigma_1}(\sigma) = \sum_{\sigma'} f_{\sigma\sigma'}\psi_{\sigma_1}(\sigma') = \sum_{\sigma'} f_{\sigma\sigma'}\delta_{\sigma_1\sigma'} = f_{\sigma\sigma_1},$$

and on substitution of this and $\psi_2(\sigma)$ the equation (40.2) is satisfied identically; this completes the proof.

[†] More precisely, we should write

$$\psi_{\sigma_1}(x, y, z; \sigma) = \psi(x, y, z)\delta_{\sigma_1\sigma}, \ldots;$$

in (40.3) the coordinate factors are omitted, being unimportant in this connection.

We must once again emphasise the distinction between the specified eigenvalue σ_1 or σ_2 of s_z and the independent variable σ.

Thus the operators acting on the wave functions of a particle with spin s can be represented in the form of $(2s+1)$-rowed matrices. In particular, we have for the operators of the spin itself

$$\hat{s}_x\psi(\sigma) = \sum_{\sigma'} (s_x)_{\sigma\sigma'}\psi(\sigma') \ldots . \tag{40.4}$$

According to what has been said above at the end of §39, the matrices $\hat{s}_x$, $\hat{s}_y$, $\hat{s}_z$ are identical with the matrices of L_x, L_y, L_z obtained in §15, where the letters L and M need only be replaced by s and σ in (15.11). This determines the spin operators.

In the important case of a spin of $\frac{1}{2}$ $(s = \frac{1}{2}, \sigma = \pm\frac{1}{2})$, these matrices have two rows, and are written in the form

$$\hat{s}_x = \tfrac{1}{2}\sigma_x, \quad \hat{s}_y = \tfrac{1}{2}\sigma_y, \quad \hat{s}_z = \tfrac{1}{2}\sigma_z, \tag{40.5}$$

where[†]

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{40.6}$$

These are called *Pauli matrices*. The matrix s_z is diagonal, as is to be expected for a matrix defined with respect to the eigenfunctions of the operator $\hat{s}_z$ itself.

§41. Spinors

Let us consider more closely the "spin" properties of wave functions. For a particle with spin zero, the wave function has only one component, which is unaltered by rotations of the coordinate system, i.e. it is a scalar.

For the wave functions of particles with non-zero spin, we must first of all consider their behaviour under rotations about the z-axis. The operator of an infinitesimal rotation through an angle $\delta\phi$ about the z-axis can be written in terms of the angular momentum operator

---

[†] In the tabular matrices (40.6) the rows and columns are numbered by the values of σ, the row number corresponding to the first and the column number to the second suffix of the matrix element. In the present case, these numbers are $\frac{1}{2}$ and $-\frac{1}{2}$. The action of the operator shown by (40.1) multiplies row σ of the matrix by a "column" containing the components of the wave function

$$\psi = \begin{pmatrix} \psi(\frac{1}{2}) \\ \psi(-\frac{1}{2}) \end{pmatrix}.$$

(in this case, the spin operator) as $1+i\,\delta\phi\cdot\hat{s}_z$. As a result of the rotation, the functions $\psi(\sigma)$ therefore become $\psi(\sigma)+\delta\psi(\sigma)$, where

$$\delta\psi(\sigma) = i\,\delta\phi\cdot\hat{s}_z\psi(\sigma).$$

But $\hat{s}_z$ is a diagonal matrix, and its diagonal elements are equal to the eigenvalues $s_z = \sigma$. Hence $\hat{s}_z\psi(\sigma) = \sigma\psi(\sigma)$, and

$$\delta\psi(\sigma) = i\sigma\psi(\sigma)\cdot\delta\phi.$$

Writing this as the differential equation $\mathrm{d}\psi/\mathrm{d}\phi = i\sigma\psi$ and integrating, we find the value of the function $\psi(\sigma)$ after a rotation through any finite angle ϕ; denoting this value by a prime, we have

$$\psi(\sigma)' = \psi(\sigma)\,e^{i\sigma\phi}. \tag{41.1}$$

In particular, a rotation through $\phi = 2\pi$ multiplies all the components $\psi(\sigma)$ by the same factor

$$e^{2\pi i\sigma} = (-1)^{2\sigma} = (-1)^{2s};$$

it is evident that the numbers 2σ always have the same parity as $2s$. Thus, when the coordinate system is completely rotated about the axis, the wave functions of particles with integral spin return to their original values, and those of particles with half-integral spin change sign.

The wave functions of particles with spin $\frac{1}{2}$, such as electrons, have two components, $\psi(\frac{1}{2})$ and $\psi(-\frac{1}{2})$. For convenience in later generalisations, we shall call these components ψ^1 and ψ^2 respectively (with superscripts 1 and 2):

$$\psi^1 = \psi(\tfrac{1}{2}), \quad \psi^2 = \psi(-\tfrac{1}{2}). \tag{41.2}$$

In any rotation of the coordinate system, ψ^1 and ψ^2 undergo a linear transformation:

$$\psi^{1\prime} = \alpha\psi^1+\beta\psi^2, \quad \psi^{2\prime} = \gamma\psi^1+\delta\psi^2. \tag{41.3}$$

The coefficients α, β, γ, δ are in general complex and functions of the angles of rotation. They are connected by relations which will be derived below.

Let us consider a system of two electrons (in relative motion with zero orbital angular momentum). The total spin S of the system may be 0 or 1. For $S = 0$, the system as a whole behaves like a particle with spin zero, so that its wave function must be a scalar. On the other hand, if the particles are regarded as not interacting, the wave function of the system must be expressed in terms of products of the wave functions of each particle separately (denoted by ψ and ϕ). It is easily seen that this wave function has to be constructed from the components of ψ and ϕ as the bilinear form

$$\frac{1}{\sqrt{2}}(\psi^1\phi^2-\psi^2\phi^1), \tag{41.4}$$

which is antisymmetric in the superscripts 1 and 2: a simple calculation using (41.3) gives

$$\frac{1}{\sqrt{2}}(\psi^{1'}\phi^{2'}-\psi^{2'}\phi^{1'}) = (\alpha\delta-\beta\gamma)\frac{1}{\sqrt{2}}(\psi^1\phi^2-\psi^2\phi^1),$$

i.e. the quantity (41.4) is transformed into itself when the coordinate system is rotated. It must therefore be a scalar. Hence we have

$$\alpha\delta-\beta\gamma = 1. \tag{41.5}$$

This is one of the required relations.

The expression

$$|\psi^1|^2+|\psi^2|^2 = \psi^1\psi^{1*}+\psi^2\psi^{2*},$$

which gives the probability of finding an electron at a particular point in space, must evidently be a scalar also. Comparison of this with the scalar (41.4) shows that the components ψ^{1*} and ψ^{2*} of the complex conjugate wave function to ψ^1, ψ^2 must be transformed as ψ^2 and $-\psi^1$ respectively:

$$\psi^{1*\prime} = \delta\psi^{1*}-\gamma\psi^{2*}, \quad \psi^{2*\prime} = -\beta\psi^{1*}+\alpha\psi^{2*}.$$

On the other hand, by taking the complex conjugate equations to (41.3),

$$\psi^{1*\prime} = \alpha^*\psi^{1*}+\beta^*\psi^{2*}, \quad \psi^{2*\prime} = \gamma^*\psi^{1*}+\delta^*\psi^{2*},$$

and comparing them with the above, we find that the coefficients $\alpha, \beta, \gamma, \delta$ are related also by

$$\alpha = \delta^*, \quad \beta = -\gamma^*. \tag{41.6}$$

By virtue of the relations (41.5), (41.6), the four complex quantities $\alpha, \beta, \gamma, \delta$ actually contain only three independent real parameters, corresponding to the three angles which define a rotation of a three-dimensional system of coordinates.

A two component quantity $\psi = \begin{pmatrix}\psi^1\\ \psi^2\end{pmatrix}$ that is transformed in accordance with (41.3) when the coordinate system is rotated is called a *spinor of rank one*, or simply a *spinor*. Thus the wave function of a particle with spin $\frac{1}{2}$ is a spinor.

Let us now consider the states of a system of two electrons having spin $S = 1$. Its wave function must have three components, corresponding to spin projections $+1$, 0, and -1. They are expressions constructed from products of components of the spinors ψ and ϕ that are symmetric with respect to the indices and are mutually transformed in accordance with (41.3):

$$\psi^1\phi^1, \quad \frac{1}{\sqrt{2}}(\psi^1\phi^2+\psi^2\phi^1), \quad \psi^2\phi^2. \tag{41.7}$$

The projection σ of the total spin of the system is equal to the sum of those of the spins of the two electrons. Thus the correspondence of the functions (41.7) to the values of σ is evident from the significance of the spinor indices 1 and 2, which show the values of the spin projections for the two electrons separately. The first function has two indices 1 and therefore corresponds to $\sigma = \frac{1}{2}+\frac{1}{2} = 1$; the second has one 1 and one 2, so that $\sigma = \frac{1}{2}-\frac{1}{2} = 0$; the third, with two indices 2, gives $\sigma = -\frac{1}{2} - \frac{1}{2} = -1$.

The spin properties of the wave functions, being essentially their properties with respect to rotations of the coordinates, are of course the same for one particle with spin 1 and for a system of particles with total spin 1. The result (41.7) is therefore more generally valid: the wave function of any particle with spin 2 is what is called a *symmetrical* *spinor of rank 2.* A spinor of rank 2 is a set of four quantities ψ^{11}, ψ^{22}, ψ^{12}, ψ^{21} that are transformed by rotations of the coordinates as products of the corresponding components of two spinors of rank 1 (but, of course, they need not actually be such products).[†] For a symmetrical spinor of rank 2, $\psi^{12} = \psi^{21}$, and it therefore has only three independent components.[‡] Their correspondence with the components of the wave function $\psi(\sigma)$ is given by the formulae

$$\psi(1) = \psi^{11}, \quad \psi(0) = \sqrt{2}\psi^{12}, \quad \psi(-1) = \psi^{22}. \tag{41.8}$$

The wave function of a particle with spin 1 can also be represented as a three-dimensional vector $\boldsymbol{\psi}$. This is evident from the fact that a three-dimensional vector is a set of likewise three quantities that are transformed into linear combinations of one another when the coordinate system is rotated. The correspondence between the components of the symmetrical spinor of rank 2 and the components of a vector is shown by the formulae

$$\psi^{11} = -(\psi_x - i\psi_y), \quad \psi^{22} = \psi_x + i\psi_y, \quad \psi^{12} = \psi_z. \tag{41.9}$$

These signify that the spinor components on the left-hand sides are transformed in the same manner as the combinations of vector components on the right-hand sides. The correctness of this assignment may be seen from the example of a rotation about the z-axis, for which the spinor transformation law is given by (41.1).[‖] On the other hand,

---

[†] Similarly, a tensor of rank 2 is a set of quantities that are transformed as the products of vector components.

[‡] An *anti*symmetrical spinor of rank 2 has only one independent component ($\psi^{11} = \psi^{22} = 0$, $\psi^{12} = -\psi^{21}$). Its properties are the same as those of the quantity (41.4) considered above. It is consequently a scalar.

[‖] According to (41.1) and (41.2),

$$\psi^{1'} = e^{i\phi/2}\psi^{1}, \quad \psi^{2'} = e^{-i\phi/2}\psi^{2},$$

where $\psi^{1'}$ and $\psi^{2'}$ are the components of the spinor in a coordinate system rotated through an angle ϕ about the z-axis relative to the original system. For the components of the spinor of rank 2, we therefore have

$$\psi^{11'} = e^{i\phi}\psi^{11}, \quad \psi^{12'} = \psi^{12}, \quad \psi^{22'} = e^{-i\phi}\psi^{22}.$$

Similar formulae relate the components of the vector $\psi_x - i\psi_y$, ψ_z, $\psi_x + i\psi_y$ in the two systems of coordinates.

from the well-known law of transformation of vector components under any rotation of the coordinates, we can find by comparison with (41.9) the general law of transformation of spinors (i.e. the dependence of the transformation coefficients (41.3) on the angles of rotation), but we shall not pause to do so here.

Lastly, in the general case of a particle with arbitrary spin, the wave function is a spinor of rank $2s$ symmetrical with respect to all its indices. It is easy to see that the number of independent components of such a spinor is equal to $2s+1$, as it should be: since the order of the indices in a symmetrical spinor has no significance, the only different components will be those whose indices include $2s$ ones and 0 twos, $2s-1$ ones and 1 two, and so on up to 0 ones and $2s$ twos.[†]

§42. Polarisation of electrons

An important property specific to particles with spin $\frac{1}{2}$ (which we shall refer to as electrons) is that, if the state of the electron is described by a wave function, then there is a direction in space along which the spin projection has the definite value $s_z = \frac{1}{2}$. (This direction may be called the direction of *polarisation* of the electron, and an electron in such a state may be said to be *completely polarised.*) By a suitable choice of the direction of the z-axis, we can always cause one component (e.g. ψ^2) of a given spinor $\psi = \begin{pmatrix}\psi^1\\ \psi^2\end{pmatrix}$, the wave function of a particle with spin $\frac{1}{2}$, to vanish. This is evident from the fact that a direction in space is determined by two quantities (for example, two angles in spherical polar coordinates), i.e. the number of disposable parameters is just equal to the number of quantities (the real and imaginary parts of the complex ψ^2) which it is desired to make zero. If $\psi^2 = 0$, the probability of the eigenvalue $s_z = -\frac{1}{2}$ is zero. For a particle with

---

[†] In mathematical terms, the symmetrical spinors of rank 1, 2, 3, ... comprise all the irreducible representations of the rotation group (see the second footnote to §18). The dimensions of these representations are $2s+1$, taking all the values 1, 2, 3, ... when $s = 0, \frac{1}{2}, 1, \ldots$. The representations given by the eigenfunctions ψ_{LM} of the orbital angular momentum (§18) are a particular case corresponding to dimensions 1, 3, 5, ...

$s > \frac{1}{2}$, it would be impossible to make all except one component of the wave function vanish in this way, since there are too many of them.

Let the z-axis be taken in the direction of polarisation of the electron. This will obviously also be the direction of the mean spin vector $\bar{\mathbf{s}}$, and its magnitude is $\frac{1}{2}$. Let us determine the probabilities $w_{\pm}$ of the values $s_{z'} = \pm\frac{1}{2}$ of the spin projection on another direction (the z'-axis) at an angle θ to the z-axis. By taking the component of $\bar{\mathbf{s}}$ along the z'-axis, we find that the mean value of the spin along this axis is $\bar{s}_{z'} = \frac{1}{2}\cos\theta$. From the definition of the probabilities $w_{\pm}$,

$$\bar{s}_{z'} = \tfrac{1}{2}(w_{+} - w_{-}).$$

Since $w_{+} + w_{-} = 1$, it follows that

$$w_{+} = \cos^2 \tfrac{1}{2}\theta, \quad w_{-} = \sin^2 \tfrac{1}{2}\theta. \tag{42.1}$$

As well as the completely polarised states of the electron, there are also states which may be said to be *partially polarised*. These are described (as regards their spin properties) not by wave functions but only by density matrices, i.e. they are mixed states (with respect to spin); similar concepts for states of orbital motion of particles have been defined in §7.

A natural means of describing such states appears if we first consider the determination of the mean spin vector in a pure state (state of complete polarisation). From the definition of the operators of physical quantities, we have[†] for a state with wave function ψ

$$\bar{\mathbf{s}} = \sum_{\alpha} \psi^{\alpha *}(\hat{\mathbf{s}}\psi^{\alpha}), \tag{42.2}$$

where the summation over the spin variable σ is represented as a summation over spinor components; in this section the letters α and β denote spinor indices taking the values 1 and 2. The bold-face letter

---

[†] In this section, as in §§40 and 41, we are not concerned with the coordinate dependence of the wave functions, and the spatial integration is therefore omitted from (42.2). The spinor ψ is then assumed normalised by the condition

$$|\psi^1|^2 + |\psi^2|^2 = 1.$$

$\boldsymbol{\sigma}$ will denote a "matrix vector" whose components are the Pauli matrices σ_x, σ_y, σ_z. According to (40.1), the effect of the spin operator $\hat{\mathbf{s}} = \frac{1}{2}\boldsymbol{\sigma}$ is the transformation

$$\hat{\mathbf{s}}\psi^\alpha = \tfrac{1}{2}\sum_\beta \boldsymbol{\sigma}^{\alpha\beta}\psi^\beta,$$

where the $\boldsymbol{\sigma}^{\alpha\beta}$ are matrix elements. We can therefore write the expression (42.2) as

$$\bar{\mathbf{s}} = \tfrac{1}{2}\sum_{\alpha,\beta} \varrho^{\beta\alpha}\boldsymbol{\sigma}^{\alpha\beta}, \tag{42.3}$$

where

$$\varrho^{\beta\alpha} = \psi^\beta\psi^{\alpha *}. \tag{42.4}$$

Evidently

$$(\varrho^{\alpha\beta})^* = \varrho^{\beta\alpha}, \tag{42.5}$$

and the normalisation condition for the wave functions gives

$$\varrho^{11} + \varrho^{22} = 1. \tag{42.6}$$

In the general case of partial polarisation, the state of the electron is described by the *polarisation density matrix*

$$\varrho^{\alpha\beta} = \begin{pmatrix} \varrho^{11} & \varrho^{12} \\ \varrho^{21} & \varrho^{22} \end{pmatrix},$$

which satisfies the conditions (42.5), (42.6) and determines $\bar{\mathbf{s}}$ according to (42.3); unlike those for a pure state, however, the elements of this matrix are not products (42.4). The magnitude of the vector $\bar{\mathbf{s}}$ can have values between 0 and $\frac{1}{2}$. The value $\frac{1}{2}$ corresponds to complete polarisation, and the value 0 to the opposite case of an unpolarised state.

The four complex quantities $\varrho^{\alpha\beta}$ are equivalent to eight real parameters, but owing to the five relations (42.5), (42.6) only three of them are independent. The real vector $\bar{\mathbf{s}}$ likewise has three components. It is therefore clear that each set uniquely determines the other. Thus the polarisation state of a particle with spin $\frac{1}{2}$ is entirely determined by specifying the mean spin vector.

The mean value of the z-component of the spin is

$$\bar{s}_z = \tfrac{1}{2}\sum_{\alpha,\beta} \sigma_z{}^{\alpha\beta}\varrho^{\beta\alpha} = \tfrac{1}{2}(\varrho^{11} - \varrho^{22}).$$

Hence we see that ϱ^{11} and ϱ^{22} are the probabilities of the eigenvalues $s_z = \frac{1}{2}$ and $s_z = -\frac{1}{2}$. The quantity ϱ^{12} is related to the mean values of s_x and s_y. Using the matrices σ_x and σ_y from (40.6), we can easily prove that

$$\varrho^{12} = \bar{s}_x - i\bar{s}_y .$$

§43. A particle in a magnetic field

A particle that has a spin also has a certain "intrinsic" magnetic moment $\boldsymbol{\mu}$. The corresponding quantum-mechanical operator is proportional to the operator $\hat{\mathbf{s}}$, and can therefore be written as

$$\hat{\boldsymbol{\mu}} = \mu \hat{\mathbf{s}}/s, \tag{43.1}$$

where s is the magnitude of the particle spin and μ a constant characterising the particle. The eigenvalues of the magnetic moment component are $\mu_z = \mu\sigma/s$. Hence we see that the coefficient μ (which is usually called just the magnitude of the magnetic moment) is the maximum possible value of μ_z, reached when $\sigma = s$.

The ratio $\mu/\hbar s$ gives the ratio of the intrinsic magnetic moment and the intrinsic angular momentum of the particle (when both are along the z-axis). For the ordinary (orbital) angular momentum, this ratio is $e/2mc$ (see *Mechanics and Electrodynamics*, §66). The coefficient of proportionality between the intrinsic magnetic moment and the spin of the particle is not the same. For an electron it is $-|e|/mc$, i.e. twice the usual value; we shall see later that this value can be obtained theoretically from Dirac's relativistic wave equation. The intrinsic magnetic moment of the electron (spin $\frac{1}{2}$) is consequently $-\mu_B$, where

$$\mu_B = |e|\hbar/2mc = 0.927 \times 10^{-20} \text{ erg/gauss}. \tag{43.2}$$

This quantity is called the *Bohr magneton.*

The magnetic moment of heavy particles is customarily measured in *nuclear magnetons*, defined as $e\hbar/2m_p c$ with m_p the mass of the proton. The intrinsic magnetic moment of the proton is found by experiment to be 2.79 nuclear magnetons, the moment being parallel to the spin. The magnetic moment of the neutron is opposite to the spin, and is 1.91 nuclear magnetons.

It should be noted that the quantities $\boldsymbol{\mu}$ and $\mathbf{s}$ on the two sides of (43.1) are the same type of vector, as they should be: both are axial vectors (both being given by vector products of two polar vectors). A similar equation for the electric dipole moment $\mathbf{d}$ ($\mathbf{d} = \text{constant}\times\mathbf{s}$) would contradict the symmetry under inversion of coordinates: the relative sign of the two sides would be changed by the inversion.[†]

Let us ascertain the form of Schrödinger's equation for a particle moving in external electric and magnetic fields. In the classical theory, the Hamilton's function for a charged particle in an electromagnetic field has the form

$$H = \frac{1}{2m}\left(\mathbf{p}-\frac{e}{c}\mathbf{A}\right)^2 + e\Phi,$$

where Φ and $\mathbf{A}$ are the scalar and vector potentials of the field, and $\mathbf{p}$ is the generalised momentum of the particle (see *Mechanics and Electrodynamics*, §43). If the particle has no spin, the transition to quantum mechanics can be made in the usual manner; the generalised momentum must be replaced by the operator $\hat{\mathbf{p}} = -i\hbar\nabla$, and we obtain the Hamiltonian[‡]

$$\hat{H} = \frac{1}{2m}\left(\hat{\mathbf{p}}-\frac{e}{c}\mathbf{A}\right)^2 + e\Phi. \tag{43.3}$$

If the particle has a spin, this procedure does not suffice. This is because the intrinsic magnetic moment of the particle interacts directly with the magnetic field. In the classical Hamilton's function this inter-

---

[†] Such an equation (and therefore the existence of an electric moment of an elementary particle) would also contradict the symmetry under time reversal: a change in the sign of the time does not alter the electric moment, but does change the sign of the spin (as is evident, for example, from the definitions of these quantities in orbital motion, that of $\mathbf{d}$ involving only the coordinates, whereas that of the magnetic moment also involves the velocity of the particle).

[‡] The generalised momentum is here denoted by the same letter $\mathbf{p}$ as the ordinary momentum, and not by $\mathbf{P}$ as in *Mechanics and Electrodynamics*, in order to emphasise that it corresponds to the same operator.

action does not appear, since the spin itself, which is a purely quantum effect, vanishes when we pass to the limit of classical mechanics. The correct expression for the Hamiltonian is obtained by adding to (43.3) a term $-\hat{\boldsymbol{\mu}}.\mathbf{H}$, which corresponds to the energy of the magnetic moment $\boldsymbol{\mu}$ in the field $\mathbf{H}$[†]. Thus the Hamiltonian of a particle having a spin and in a magnetic field is

$$\hat{H} = \frac{1}{2m}\left(\hat{\mathbf{p}}-\frac{e}{c}\mathbf{A}\right)^2-\boldsymbol{\mu}.\mathbf{H}. \tag{43.4}$$

The equation $\hat{H}\psi = E\psi$ for the eigenvalues of this operator is the required generalisation of Schrödinger's equation to the case of motion in a magnetic field. The wave function ψ in this equation is a spinor of rank $2s+1$.

§44. Motion in a uniform magnetic field

Let us determine the energy levels of an electron in a constant uniform magnetic field. We take the z-axis in the direction of the field $\mathbf{H}$, and write the vector potential of the field in the form

$$A_x = -Hy, \quad A_y = A_z = 0; \tag{44.1}$$

it is then easily verified that curl $\mathbf{A} = \mathbf{H}$. The Hamiltonian (43.4) of the electron (with charge $e = -|e|$ and magnetic moment $\mu = -\mu_B$) becomes

$$\hat{H} = \frac{1}{2m}\left(\hat{p}_x+\frac{eH}{c}y\right)^2+\frac{\hat{p}_y^2}{2m}+\frac{\hat{p}_z^2}{2m}-\frac{eH}{mc}\hat{s}_z. \tag{44.2}$$

First of all, we notice that the operator $\hat{s}_z$ commutes with the Hamiltonian, since the latter does not contain the operators of the other components of the spin. This means that the z-component of the spin is conserved, and therefore that $\hat{s}_z$ can be replaced by the eigenvalues $s_z = \sigma$. Then the spin dependence of the wave function

---

[†] There should be no misunderstanding here caused by the use of the same letter for the field and the Hamiltonian, since the latter always has a circumflex over it.

becomes unimportant, and ψ in Schrödinger's equation can be taken as the ordinary coordinate function. For this function we have the equation

$$\frac{1}{2m}\left[\left(\hat{p}_x+\frac{eH}{c}y\right)^2+\hat{p}_y^2+\hat{p}_z^2\right]\psi-\frac{eH}{mc}\psi = E\psi. \tag{44.3}$$

The Hamiltonian (44.2) does not contain the coordinates x and z explicitly. The operators $\hat{p}_x$ and $\hat{p}_z$ (of differentiation with respect to x and z) therefore also commute with the Hamiltonian, i.e. the x- and z-components of the generalised momentum are conserved. We accordingly seek ψ in the form

$$\psi = e^{(i/\hbar)(p_x x+p_z z)}\chi(y). \tag{44.4}$$

The eigenvalues p_x and p_z take all values from $-\infty$ to $+\infty$. Since $A_z = 0$, the z-component p_z of the generalised momentum is equal to the ordinary momentum component: $p_z = mv_z$. Thus the velocity of the electron in the direction of the field can take any value; we can say that the motion along the field is "not quantised".

Substituting (44.4) in (44.3), we obtain the following equation for the function χ:

$$\chi''+\frac{2m}{\hbar^2}\left[E-\omega_H\sigma-\frac{p_z^2}{2m}-\tfrac{1}{2}m\omega_H^2(y-y_0)^2\right]\chi = 0,$$

with the notation $y_0 = -cp_x/eH$ and

$$\omega_H = |e|\,H/mc. \tag{44.5}$$

This equation is formally identical with Schrödinger's equation (25.6) for a linear oscillator, oscillating with frequency ω_H about the point $y = y_0$. Hence we can conclude immediately that the constant $E-\omega_H\sigma-p_z^2/2m$, which takes the part of the oscillator energy, can have the values $(n+\frac{1}{2})\hbar\omega_H$, where n is any integer.

Thus we obtain the following expression for the energy levels of an electron in a uniform magnetic field:

$$E = (n+\tfrac{1}{2}+\sigma)\hbar\omega_H+p_z^2/2m. \tag{44.6}$$

The first term in (44.6) gives the discrete energy levels, corresponding to motion in a plane perpendicular to the field; they are called *Landau levels*.[†]

[†] This problem was first investigated by L. D. Landau (1930) in connection with the study of electron diamagnetism in metals.

CHAPTER 6

IDENTITY OF PARTICLES

§45. The principle of indistinguishability of similar particles

In classical mechanics, identical particles (electrons, say) do not lose their "individuality", despite the identity of their physical properties. For we can imagine the particles in a physical system at some instant to be "numbered", and follow the subsequent motion of each of these in its path; then at any instant the particles can be identified.

In quantum mechanics the situation is entirely different. We have already mentioned several times that, by virtue of the uncertainty principle, the concept of the path of an electron ceases to have any meaning. If the position of an electron is exactly known at a given instant, its coordinates have no definite values even at the next instant. Hence, by localising and numbering the electrons at some instant, we make no progress towards identifying them at subsequent instants; if we localise one of the electrons, at some other instant, at some point in space, we cannot say which of the electrons has arrived at this point.

Thus, in quantum mechanics, there is in principle no possibility of separately following each of a number of similar particles and thereby distinguishing them. We may say that, in quantum mechanics, identical particles entirely lose their "individuality". The identity of the particles with respect to their physical properties is here very far-reaching: it results in the complete indistinguishability of the particles.

This principle of the *indistinguishability of similar particles*, as it is called, plays a fundamental part in the quantum theory of systems composed of identical particles. Let us start by considering a system of only two particles. Because of the identity of the particles, the states of the system obtained from each other by merely interchanging the two particles must be completely equivalent physically. This means that, as a result of this interchange, the wave function of the system can change only by an unimportant phase factor. Let $\psi(\xi_1, \xi_2)$ be the wave function of the system, ξ_1 and ξ_2 conventionally denoting the three coordinates and the spin projection for each particle. Then we must have

$$\psi(\xi_1, \xi_2) = e^{i\alpha}\psi(\xi_2, \xi_1),$$

where α is some real constant. By repeating the interchange, we return to the original state, while the function ψ is multiplied by $e^{2i\alpha}$. Hence it follows that $e^{2i\alpha} = 1$, or $e^{i\alpha} = \pm 1$. Thus

$$\psi(\xi_1, \xi_2) = \pm\psi(\xi_2, \xi_1).$$

We thus reach the result that there are only two possibilities: the wave function is either *symmetrical* (i.e. it is unchanged when the particles are interchanged) or *antisymmetrical* (i.e. it changes sign when this interchange is made). It is obvious that the wave functions of all the states of a given system must have the same symmetry; otherwise, the wave function of a state which was a superposition of states of different symmetry would be neither symmetrical nor antisymmetrical.

This result can be immediately generalised to systems consisting of any number of identical particles. For it is clear from the identity of the particles that, if any pair of them has the property of being described by, say, symmetrical wave functions, any other pair of such particles has the same property. Hence the wave function of identical particles must either be unchanged when any pair of particles are interchanged (and hence when the particles are permuted in any manner), or change sign when any pair are interchanged. In the first case we speak of a *symmetrical* wave function, and in the second case of an *antisymmetrical* one.

The property of being described by symmetrical or antisymmetrical wave functions depends on the nature of the particles. Particles described by antisymmetrical functions are said to obey *Fermi–Dirac statistics* (or to be *fermions*), while those which are described by symmetrical functions are said to obey *Bose–Einstein statistics* (or to be *bosons*).†

We shall see later (§87) that, according to the laws of relativistic quantum mechanics, the statistics obeyed by particles is uniquely related to their spin: particles with half-integral spin are fermions, and those with integral spin are bosons.

The statistics of complex particles is determined by the parity of the number of elementary fermions entering into their composition. For an interchange of two identical complex particles is equivalent to the simultaneous interchange of several pairs of identical elementary particles. The interchange of bosons does not change the wave function, while the interchange of fermions changes its sign. Hence complex particles containing an odd number of elementary fermions obey Fermi statistics, while those containing an even number obey Bose statistics. This result is, of course, in agreement with the above rule, since a complex particle has an integral or a half-integral spin according as the number of particles with half-integral spin entering into its composition is even or odd.

Thus atomic nuclei of odd atomic weight (i.e. containing an odd number of neutrons and protons) obey Fermi statistics, and those of even atomic weight obey Bose statistics. For atoms, which contain both nuclei and electrons, the statistics is evidently determined by the parity of the sum of the atomic weight and the atomic number.

Let us consider a system composed of N identical particles, whose mutual interaction can be neglected. Let ψ_1, ψ_2, ... be the wave func-

---

† This terminology refers to the statistics which describes a perfect gas composed of particles with antisymmetrical and symmetrical wave functions respectively. In actual fact we are concerned here not only with a different statistics, but essentially with a different mechanics. Fermi statistics was proposed by E. Fermi for electrons in 1926, and its relation to quantum mechanics was elucidated by P. A. M. Dirac (1926). Bose statistics was proposed by S. N. Bose for light quanta, and generalised by A. Einstein (1924).

tions of the various stationary states which each of the particles separately may occupy. The state of the system as a whole can be defined by giving the numbers of the states which the individual particles occupy. The question arises how the wave function ψ of the whole system should be constructed from the functions $\psi_1, \psi_2, \dots$.

Let $p_1, p_2, \dots, p_N$ be the numbers of the states occupied by the individual particles (some of these numbers may be the same). For a system of bosons, the wave function $\psi(\xi_1, \xi_2, \dots, \xi_N)$ is given by a sum of products of the form

$$\psi_{p_1}(\xi_1)\,\psi_{p_2}(\xi_2)\dots\psi_{p_N}(\xi_N), \tag{45.1}$$

with all possible permutations of the different suffixes $p_1, p_2, \dots$; this sum clearly possesses the required symmetry property. Thus, for example, for a system of two particles in different states ($p_1 \neq p_2$)

$$\psi(\xi_1, \xi_2) = [\psi_{p_1}(\xi_1)\,\psi_{p_2}(\xi_2) + \psi_{p_1}(\xi_2)\,\psi_{p_2}(\xi_1)]/\sqrt{2}. \tag{45.2}$$

The factor $1/\sqrt{2}$ is introduced for normalisation purposes; all the functions $\psi_1, \psi_2, \dots$ are orthogonal and are supposed normalised. In the general case of a system containing an arbitrary number N of particles, the normalised wave function is

$$\psi = \left(\frac{N_1!\,N_2!\dots}{N!}\right)^{1/2} \Sigma\psi_{p_1}(\xi_1)\,\psi_{p_2}(\xi_2)\dots\psi_{p_N}(\xi_N), \tag{45.3}$$

where the sum is taken over all permutations of the different suffixes $p_1, p_2, \dots, p_N$, and the numbers N_i show how many of these suffixes have the same value i (with $\Sigma N_i = N$). In the integration of $|\psi|^2$ over $\xi_1, \xi_2, \dots, \xi_N$, all terms vanish except the squared modulus of each term[†]; since the total number of terms in the sum (45.3) is evidently

$$N!/N_1!\,N_2!\dots,$$

the normalisation factor in (45.3) is obtained.

[†] The integration over ξ is conventionally understood in §§45–47 as including integration over the coordinates and summation over σ.

For a system of fermions, the wave function ψ is an antisymmetrical combination of the products (45.1). For a system composed of two particles we have

$$\psi = [\psi_{p_1}(\xi_1)\psi_{p_2}(\xi_2)-\psi_{p_1}(\xi_2)\psi_{p_2}(\xi_1)]/\sqrt{2}. \qquad (45.4)$$

For the general case of N particles, the wave function can be written in the form of a determinant

$$\Psi = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_{p_1}(\xi_1) & \psi_{p_1}(\xi_2) & \dots & \psi_{p_1}(\xi_N) \\ \psi_{p_2}(\xi_1) & \psi_{p_2}(\xi_2) & \dots & \psi_{p_2}(\xi_N) \\ \dots & \dots & \dots & \dots \\ \psi_{p_N}(\xi_1) & \psi_{p_N}(\xi_2) & \dots & \psi_{p_N}(\xi_N) \end{vmatrix}. \qquad (45.5)$$

An interchange of two particles corresponds to an interchange of two columns of the determinant, as a result of which the latter changes sign.

The following important result is a consequence of the expression (45.5). If among the numbers $p_1, p_2, \dots$ any two are the same, two rows of the determinant are the same, and it therefore vanishes identically. It will be different from zero only when all the numbers $p_1, p_2, \dots$ are different. Thus, in a system consisting of identical fermions, no two (or more) particles can be in the same state at the same time. This is called *Pauli's principle*; it was established by W. Pauli in 1925.

§46. Exchange interaction

The fact that Schrödinger's equation does not take account of the spin of particles does not invalidate this equation or the results obtained by means of it. This is because the electrical interaction of the particles does not depend on their spins.[†] Mathematically, this means that the Hamiltonian of a system of electrically interacting particles (in the absence of a magnetic field) does not contain the spin operators, and hence, when it is applied to the wave function, it has no effect on

---

[†] This is true only so long as we consider the non-relativistic approximation. When relativistic effects are taken into account, the interaction of charged particles does depend on their spin.

the spin variables. Hence Schrödinger's equation is actually satisfied by each component of the wave function, i.e. the wave function of the system of particles can be written in the form of a product

$$\psi(\xi_1, \xi_2, \ldots) = \chi(\sigma_1, \sigma_2, \ldots)\,\phi(\mathbf{r}_1, \mathbf{r}_2, \ldots), \tag{46.1}$$

where the function ϕ depends only on the coordinates of the particles and the function χ only on their spins. We call the former a *coordinate* or *orbital* wave function, and the latter a *spin* wave function. Schrödinger's equation essentially determines only the coordinate function ϕ, the function χ remaining arbitrary. In any instance where we are not interested in the actual spin of the particles, we can therefore use Schrödinger's equation and regard as the wave function the coordinate function alone, as we have done hitherto.

However, despite the fact that the electrical interaction of the particles is independent of their spin, there is a peculiar dependence of the energy of the system on its total spin, arising ultimately from the principle of indistinguishability of similar particles.

Let us consider a system consisting of only two identical particles. By solving Schrödinger's equation we find a series of energy levels, to each of which there corresponds a definite symmetrical or antisymmetrical coordinate wave function $\phi(\mathbf{r}_1, \mathbf{r}_2)$. For, by virtue of the identity of the particles, the Hamiltonian (and therefore the Schrödinger's equation) of the system is invariant with respect to interchange of the particles. If the energy levels are not degenerate, the function $\phi(\mathbf{r}_1, \mathbf{r}_2)$ can change only by a constant factor when the coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$ are interchanged; repeating this interchange, we see that this factor can only be[†] ± 1.

Let us first suppose that the particles have zero spin. The spin factor for such particles is absent altogether, and the wave function reduces to the coordinate function $\phi(\mathbf{r}_1, \mathbf{r}_2)$, which must be symmetrical (since particles with zero spin obey Bose statistics). Thus not all the energy levels obtained by a formal solution of Schrödinger's equation can

---

[†] When there is degeneracy we can always choose linear combinations of the functions belonging to a given level, such that this condition is again satisfied.

actually exist; those to which antisymmetrical functions ϕ correspond are not possible for the system under consideration.

The interchange of two similar particles is equivalent to the operation of inversion of the coordinate system (the origin being taken to bisect the line joining the two particles). On the other hand, the result of inversion is to multiply the wave function ϕ by $(-1)^l$, where l is the orbital angular momentum of the relative motion of the two particles (see §19). By comparing these considerations with those given above, we conclude that a system of two identical particles with zero spin can have only an even orbital angular momentum.

Next, let us suppose that the system consists of two particles with spin $\frac{1}{2}$ (say, electrons). Then the complete wave function of the system (i.e. the product of the function $\phi(\mathbf{r}_1, \mathbf{r}_2)$ and the spin function $\chi(\sigma_1, \sigma_2)$) must certainly be antisymmetrical with respect to an interchange of the two electrons. Hence, if the coordinate function is symmetrical, the spin function must be antisymmetrical, and vice versa. We shall write the spin function in spinor form, i.e. as a spinor $\chi^{\alpha\beta}$ of rank two, each of whose indices corresponds to the spin of one of the electrons. A symmetrical spinor ($\chi^{\alpha\beta} = \chi^{\beta\alpha}$) corresponds to a function symmetrical with respect to the spins of the two particles, and an antisymmetrical spinor ($\chi^{\alpha\beta} = -\chi^{\beta\alpha}$) to an antisymmetrical function. We know, however, that a symmetrical spinor of rank two describes a system with total spin unity, while an antisymmetrical spinor reduces to a scalar, corresponding to zero spin.

Thus we reach the following conclusion. The energy levels to which there correspond symmetrical solutions $\phi(\mathbf{r}_1, \mathbf{r}_2)$ of Schrödinger's equation can actually occur when the total spin of the system is zero, i.e. when the spins of the two electrons are "antiparallel", giving a sum of zero. The values of the energy belonging to antisymmetrical functions $\phi(\mathbf{r}_1, \mathbf{r}_2)$, on the other hand, require a value of unity for the total spin, i.e. the spins of the two electrons must be "parallel".

In other words, the possible values of the energy of a system of electrons depend on their total spin. For this reason we can speak of a peculiar interaction of the particles which results in this dependence. This is called *exchange interaction*. It is a purely quantum effect, which entirely vanishes (like the spin itself) in the passage to the limit of classical mechanics.

§47. Second quantisation. The case of Bose statistics

In the theory of systems consisting of a very large number of identical particles there is a widely used method of considering the problem, known as *second quantisation.* This method is necessary in relativistic theory, where we have to deal with systems in which the number of particles is itself variable.[†]

Let $\psi_1(\xi)$, $\psi_2(\xi)$, ... be some complete set of orthogonal and normalised wave functions of stationary states of a single particle. These may be taken as plane waves, i.e. the wave functions of a free particle having definite values of the momentum (and spin projection); in order to make the spectrum of states discrete we shall consider the motion of particles in a large but finite volume Ω, as described at the end of §27.

In a system of free particles, the particle momenta are separately conserved. The *occupation numbers* of the states are therefore also conserved, i.e. the numbers N_1, N_2, ... which show how many particles are in each of the states ψ_1, ψ_2, ... In a system of interacting particles, the momentum of each particle is not conserved, and so the occupation numbers are not conserved. For such a system we can consider only the probability distribution of the various values of the occupation numbers. Let us seek to construct a mathematical formalism in which the occupation numbers (and not the coordinates and spin projections of the particles) play the part of independent variables.

In this formalism, the state of the system is described by what is called a wave function in occupation-number space, denoted by $\Phi(N_1, N_2 \ldots; t)$ in order to emphasise the difference from the ordinary coordinate-spin wave function $\Psi(\xi_1, \xi_2, \ldots, \xi_N; t)$. The squared modulus $|\Phi|^2$ determines the probability of the various values of the numbers $N_1, N_2, \ldots$.

[†] The method of second quantisation was developed by P. A. M. Dirac (1927) for photons in radiation theory, and later extended to fermions by E. Wigner and P. Jordan (1928).

In accordance with this choice of the independent variables, the operators of the various physical quantities (including the Hamiltonian of the system) must be formulated in terms of their action on functions of the occupation numbers. Such a formulation can be obtained on the basis of the usual matrix representation of operators. The operator matrix elements must be considered in relation to the wave functions of the stationary states of a system of non-interacting particles. Since these states can be described by specifying definite values of the occupation numbers, this will also show the nature of the action of the operators on these variables.

Let us first consider systems of particles obeying Bose statistics. Let $\hat{f}_a^{(1)}$ be the operator of some quantity pertaining to the ath particle, i.e. acting only on functions of ξ_a. We introduce the operator

$$\hat{F}^{(1)} = \sum_a \hat{f}_a^{(1)}, \tag{47.1}$$

which is symmetrical with respect to all the particles (the summation being over all particles) and determine its matrix elements with respect to the wave functions (45.3). First of all, it is easy to see that the matrix elements will be different from zero only for transitions which leave the numbers $N_1, N_2, \ldots$ unchanged (diagonal elements) and for transitions where one of these numbers is increased, and another decreased, by unity. For, since each of the operators $\hat{f}_a^{(1)}$ acts on only one function in the product $\psi_{p_1}(\xi_1)\psi_{p_2}(\xi_2)\ldots\psi_{p_N}(\xi_N)$, its matrix elements can be different from zero only for transitions whereby the state of a single particle is changed; this, however, means that the number of particles in one state is diminished by unity, while the number in another state is correspondingly increased. The calculation of these matrix elements is in principle very simple; it is easier to do it oneself than to follow an account of it. Hence we shall give only the result of the calculation. The non-diagonal elements are

$$\langle N_i, N_k-1 \,|\, F^{(1)} \,|\, N_i-1, N_k\rangle = f_{ik}^{(1)}\sqrt{(N_i N_k)}. \tag{47.2}$$

We shall indicate only those suffixes with respect to which the matrix element is non-diagonal, omitting the remainder for brevity. Here $f_{ik}^{(1)}$ is the matrix element

$$f_{ik}^{(1)} = \int \psi_i^*(\xi) \hat{f}^{(1)} \psi_k(\xi)\, d\xi. \tag{47.3}$$

Since the operators $\hat{f}_a^{(1)}$ differ only in the naming of the variable on which they act, the integrals (47.3) are independent of a, and this suffix is therefore omitted. The diagonal matrix elements of $F^{(1)}$ are the mean values of the quantity $F^{(1)}$ in the states $\psi_{N_1 N_2 \ldots}$. Calculation gives

$$\overline{F^{(1)}} = \sum_i f_{ii}^{(1)} N_i. \tag{47.4}$$

We now introduce the operators $\hat{a}_i$, which play a leading part in the method of second quantisation; they act, not on functions of the coordinates, but on functions of the occupation numbers. By definition, the operator $\hat{a}_i$ acting on the function $\Phi(N_1, N_2, \ldots)$ decreases the value of the variable N_i by unity, and at the same time it multiplies the function by $\sqrt{N_i}$:

$$\hat{a}_i \Phi(N_1, N_2, \ldots, N_i, \ldots) = \sqrt{(N_i)}\, \Phi(N_1, N_2, \ldots, N_i - 1, \ldots). \tag{47.5}$$

We can say that the operator $\hat{a}_i$ diminishes by one the number of particles in the ith state; it is therefore called a particle *annihilation operator*. It can be represented in the form of a matrix whose only non-zero element is

$$\langle N_i - 1 | a_i | N_i \rangle = \sqrt{N_i}. \tag{47.6}$$

The operator $\hat{a}_i^+$, which is the Hermitian conjugate of $\hat{a}_i$, is, by definition (see (11.9)), represented by a matrix whose only non-zero element is

$$\langle N_i | a_i^+ | N_i - 1 \rangle = \langle N_i - 1 | a_i | N_i \rangle^* = \sqrt{N_i}. \tag{47.7}$$

This means that, when acting on the function $\Phi(N_1, N_2, \ldots)$, it increases the number N_i by unity:

$$\hat{a}_i^+ \Phi(N_1, N_2, \ldots, N_i, \ldots) = \sqrt{(N_i + 1)}\, \Phi(N_1, N_2, \ldots, N_i + 1, \ldots). \tag{47.8}$$

In other words, the operator $\hat{a}_i^+$ increases by one the number of particles in the ith state, and is therefore called a particle *creation operator*.

The product of operators $\hat{a}_i^+\hat{a}_i$, acting on the wave function, must multiply it by a constant simply, leaving unchanged all the variables $N_1, N_2, \ldots$: the operator $\hat{a}_i$ diminishes N_i by unity, and $\hat{a}_i^+$ then restores it to its original value. Direct multiplication of the matrices (47.6) and (47.7) shows that $\hat{a}_i^+\hat{a}_i$ is represented, as we should expect, by a diagonal matrix whose diagonal elements are N_i. We can write

$$\hat{a}_i^+\hat{a}_i = N_i. \qquad (47.9)$$

Similarly, we find that

$$\hat{a}_i\hat{a}_i^+ = N_i+1. \qquad (47.10)$$

The difference of these equations gives the commutation rule for the operators $\hat{a}_i$ and $\hat{a}_i^+$:

$$\hat{a}_i\hat{a}_i^+-\hat{a}_i^+\hat{a}_i = 1. \qquad (47.11)$$

The operators with i and k different act on different variables (N_i and N_k), and of course commute:

$$\hat{a}_i\hat{a}_k-\hat{a}_k\hat{a}_i = 0, \quad \hat{a}_i\hat{a}_k^+-\hat{a}_k^+\hat{a}_i = 0 \qquad (i \neq k). \qquad (47.12)$$

From the above properties of the operators $\hat{a}_i$, $\hat{a}_i^+$ it is easy to see that the operator

$$\hat{F}^{(1)} = \sum_{i,k} f_{ik}^{(1)}\hat{a}_i^+\hat{a}_k \qquad (47.13)$$

is the same as the operator (47.1). For all the matrix elements calculated from (47.6), (47.7) are the same as the elements (47.2). This is a very important result. In formula (47.13), the quantities $f_{ik}^{(1)}$ are simply numbers. Thus we have been able to express an ordinary operator, acting on functions of the coordinates, in the form of an operator acting on functions of new variables, the occupation numbers N_i.

The result which we have obtained is easily generalised to operators of other forms. Let

$$\hat{F}^{(2)} = \sum_{a>b} \hat{f}_{ab}^{(2)}, \qquad (47.14)$$

where $\hat{f}_{ab}^{(2)}$ is the operator of a physical quantity pertaining to two particles at once, and hence acts on functions of ξ_a and ξ_b. Similar calculations show that this operator can be expressed in terms of the operators $\hat{a}_i$, $\hat{a}_i^+$ by

$$\hat{F}^{(2)} = \tfrac{1}{2} \sum_{i,k,l,m} (f^{(2)})_{lm}^{ik} \hat{a}_i^+ \hat{a}_k^+ \hat{a}_m \hat{a}_l, \tag{47.15}$$

where

$$(f^{(2)})_{lm}^{ik} = \iint \psi_i^*(\xi_1)\,\psi_k^*(\xi_2)\,\hat{f}^{(2)}\psi_l(\xi_1)\,\psi_m(\xi_2)\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2.$$

The formulae can be similarly generalised to operators of any other form symmetrical with respect to all the particles.

Using these formulae, we can express in terms of the operators $\hat{a}_i$, $\hat{a}_i^+$ also the Hamiltonian of the physical system of N interacting particles that is actually being considered. The Hamiltonian of such a system is, of course, symmetrical with respect to all the particles. For example, if the interaction in the system consists of interactions between each pair of particles, the Hamiltonian is

$$\hat{H} = \sum_a \hat{H}_a^{(1)} + \sum_{a>b} U^{(2)}(\mathbf{r}_a, \mathbf{r}_b). \tag{47.16}$$

Here $\hat{H}_a^{(1)}$ is the part of the Hamiltonian which depends on the coordinates of only one particle, i.e. the free-particle Hamiltonian:

$$\hat{H}_a^{(1)} = -(\hbar^2/2m)\triangle_a. \tag{47.17}$$

The function $U^{(2)}(\mathbf{r}_a, \mathbf{r}_b)$ is the energy of interaction of two particles. Applying formulae (47.13) and (47.15) to (47.16), we obtain

$$\hat{H} = \sum_{i,k} H_{ik}^{(1)} \hat{a}_i^+ \hat{a}_k + \tfrac{1}{2} \sum_{i,k,l,m} (U^{(2)})_{lm}^{ik} \hat{a}_i^+ \hat{a}_k^+ \hat{a}_m \hat{a}_l. \tag{47.18}$$

This gives the required expression for the Hamiltonian in the form of an operator acting on functions of the occupation numbers.

For a system of non-interacting particles, only the first term in the expression (47.18) remains:

$$\hat{H} = \sum_{i,k} H_{ik}^{(1)} \hat{a}_i^+ \hat{a}_k. \tag{47.19}$$

If the functions ψ_i are taken to be, as agreed, the eigenfunctions of the Hamiltonian $\hat{H}^{(1)}$ of a free particle, the matrix $H_{ik}^{(1)}$ is diagonal, and its diagonal elements are the eigenvalues ε_i of the energy of the particle. Thus

$$\hat{H} = \sum_i \varepsilon_i \hat{a}_i^+ \hat{a}_i; \tag{47.20}$$

replacing the operator $\hat{a}_i^+ \hat{a}_i$ by its eigenvalues (47.9), we have for the energy levels of the system the expression

$$E = \sum_i \varepsilon_i N_i, \tag{47.21}$$

a trivial result which could have been foreseen.

The formalism of second quantisation can be put in a more compact form by introducing the operators

$$\hat{\psi}(\xi) = \sum_i \psi_i(\xi)\,\hat{a}_i, \quad \hat{\psi}^+(\xi) = \sum_i \psi_i^*(\xi)\,\hat{a}_i^+, \tag{47.22}$$

where the variables ξ are regarded as parameters. By what has been said above concerning the operators $\hat{a}_i$, $\hat{a}_i^+$, it is clear that the operator $\hat{\psi}$ decreases the total number of particles in the system by one, while $\hat{\psi}^+$ increases it by one.[†]

By means of the ψ-operators, the Hamiltonian (47.18) can be written in the form

$$\hat{H} = \int \hat{\psi}^+(\xi)\,\hat{H}^{(1)}\hat{\psi}(\xi)\,\mathrm{d}\xi + \tfrac{1}{2}\iint \hat{\psi}^+(\xi)\,\psi^+(\xi')\,U^{(2)}\hat{\psi}(\xi')\,\hat{\psi}(\xi)\,\mathrm{d}\xi\,\mathrm{d}\xi'. \tag{47.23}$$

This is easily proved by direct substitution of the ψ-operators (47.22).

The operator $\hat{\psi}^+\hat{\psi}$, constructed from the ψ-operators in the same way as the product $\psi^*\psi$ giving the probability density for a particle in a state with wave function ψ, is called the *particle density operator*. The integral

$$\hat{N} = \int \hat{\psi}^+\hat{\psi}\,\mathrm{d}\xi \tag{47.24}$$

represents in the second quantisation formalism the operator of the total number of particles in the system. For, substituting the ψ-operators in the form (47.22) and using the normalisation and orthogonality of the wave functions ψ_i, we have

$$\hat{N} = \sum_i \hat{a}_i^+\hat{a}_i. \tag{47.25}$$

Each term in this sum is the operator of the number of particles in the

---

[†] Attention is drawn to the analogy between the expressions (47.22) and the expansion $\psi = \Sigma a_i\psi_i$ of an arbitrary wave function in terms of a complete set of functions. Here the expansion is, so to speak, quantised afresh. This is the reason for calling the method the *second quantisation* method.

ith state: according to (47.9), its eigenvalues are equal to the occupation numbers N_i, and the sum of all these numbers is the total number of particles in the system. For systems containing a specified number of particles, these statements are trivial, as are the properties of the Hamiltonian (47.19) of a system of free particles. We shall see, however, that their generalisation in the relativistic theory yields new results that are by no means trivial.

§48. Second quantisation. The case of Fermi statistics

The basic theory of the method of second quantisation remains wholly unchanged for systems of identical fermions, but the actual formulae for the matrix elements of quantities and for the operators $\hat{a}_i$ are naturally different.

We shall not pause here to give the details of the calculations, but merely emphasise the important points in them, where there are differences from those in §47.

The wave function $\psi_{N_1 N_2} \ldots$ now has the form (45.5). As already mentioned, the numbers $p_1, p_2, \ldots$ which number the occupied states must all be different, since otherwise the determinant would vanish. In other words, the occupation numbers N_i can take only the values 0 and 1.

Because of the antisymmetry of the function (45.5), the question of its sign arises first of all. This question did not arise in the case of Bose statistics, since, because of the symmetry of the wave function, its sign, once chosen, was preserved under all permutations of the particles. In order to make definite the sign of the function (45.5), we can agree to choose it as follows. We number successively, once and for all, all the states ψ_i. We then complete the rows of the determinant (45.5) so that $p_1 < p_2 < p_3 < \ldots < p_N$, whilst in the successive columns we have functions of the different variables in the order $\xi_1, \xi_2, \ldots, \xi_N$. The sign of the wave function will thus depend on the set of numbers $p_1, p_2, \ldots$, i.e. on all the occupation numbers.

It is then found that the signs of the matrix elements of the particle annihilation and creation operators also depend on these numbers:

the operators have to be defined as matrices having a single non-zero element

$$\langle 0_i | a_i | 1_i \rangle = \langle 1_i | a_i^+ | 0_i \rangle = (-1)^{\sum\limits_{k=1}^{i-1} N_k} . \tag{48.1}$$

By multiplying these matrices it can be shown that the products $\hat{a}_i^+ \hat{a}_i$ and $\hat{a}_i \hat{a}_i^+$ are diagonal, with

$$\hat{a}_i^+ \hat{a}_i = N_i, \quad \hat{a}_i \hat{a}_i^+ = 1 - N_i, \tag{48.2}$$

and their sum is

$$\hat{a}_i \hat{a}_i^+ + \hat{a}_i^+ \hat{a}_i = 1. \tag{48.3}$$

It should be noticed that the vanishing of the product $\hat{a}_i^+ \hat{a}_i$ when $N_i = 0$ and of $\hat{a}_i \hat{a}_i^+$ when $N_i = 1$ is to be expected. In these products, the operator on the right acts first, but a particle in the *i*th state cannot be annihilated if it does not exist ($N_i = 0$) and according to Pauli's principle an electron in the *i*th state cannot be created if this state is already occupied, i.e. if $N_i = 1$. For the same reason, it is obvious that

$$\hat{a}_i \hat{a}_i = 0, \quad \hat{a}_i^+ \hat{a}_i^+ = 0. \tag{48.4}$$

For any pair of operators with different *i* and *k*, we have

$$\hat{a}_i \hat{a}_k + \hat{a}_k \hat{a}_i = 0, \quad \hat{a}_i^+ \hat{a}_k^+ + \hat{a}_k^+ \hat{a}_i^+ = 0, \quad \hat{a}_i \hat{a}_k^+ + \hat{a}_k^+ \hat{a}_i = 0 \qquad (i \neq k); \tag{48.5}$$

that is, they all *anticommute*, the product changing sign when the factors are transposed. This difference from the case of Bose statistics is perfectly natural. In the latter case, the operators $\hat{a}_i$ and $\hat{a}_k$ were completely independent; each of the operators $\hat{a}_i$ acted only on a single variable N_i, and the result of this action did not depend on the values of the other occupation numbers. In the case of Fermi statistics, however, the result of the action of the operator $\hat{a}_i$ depends not only on the number N_i itself, but also on the occupation numbers of all the preceding states. Hence the action of the various operators $\hat{a}_i$, $\hat{a}_k$ cannot be considered independent.

When the properties of the operators $\hat{a}_i$, $\hat{a}_i^+$ have thus been established, all the remaining formulae (47.13)–(47.25) remain valid.

CHAPTER 7

THE ATOM

§49. Atomic energy levels

In the non-relativistic approximation, the stationary states of the atom are determined by Schrödinger's equation for the system of electrons, which move in the Coulomb field of the nucleus and interact electrically with one another. As we know, for a system of particles in a centrally symmetric external field the total orbital angular momentum L and the parity of the state are conserved. Hence each stationary state of the atom will be characterised by a definite value of the orbital angular momentum L and by its parity. Moreover, because of the exchange interaction described in §46, every stationary state of the atom is characterised also by a definite value of the total spin S of the electrons.

Thus, in the non-relativistic approximation, the energy levels of the atom are classified by means of the values of L and S and of the parity (the converse is of course not true: the values of these quantities by themselves do not uniquely determine the energy). Each such energy level is degenerate in accordance with the different possible directions in space of the vectors **L** and **S**. The degrees of degeneracy with respect to these directions are respectively $2L+1$ and $2S+1$. Consequently, the total degree of degeneracy of a level with given L and S is equal to the product $(2L+1)(2S+1)$.

In fact, however, there are relativistic effects in the electromagnetic interaction of electrons; these effects depend on the spins of the electrons, and will be further considered in §51. They have the result that the energy of the atom depends not only on the absolute magnitudes of the vectors **L** and **S** but also on their relative positions. Strictly speaking, when the relativistic interactions are taken into account the orbital angular momentum **L** and the spin **S** of the atom are not separately conserved. Only the total angular momentum $\mathbf{J} = \mathbf{L}+\mathbf{S}$ is conserved; this is a universal and exact law which follows from the isotropy of space relative to a closed system. For this reason, the exact energy levels of the atom must be characterised by the values J of the total angular momentum.

However, if the relativistic effects are comparatively small (as happens in many cases), they can be allowed for as a perturbation. Under the action of this perturbation, a level with given L and S is "split" into a number of distinct (though close) levels, which differ in the value of the total angular momentum J. These levels are determined (in the first approximation) by the appropriate secular equation (§33), while their wave functions (in the zeroth approximation) are definite linear combinations of the wave functions of the initial degenerate level with the given L and S. In this approximation we can therefore, as before, regard the absolute values of the orbital angular momentum and spin (but not their directions) as being conserved, and characterise the levels by the values of L and S also.

Thus, as a result of the relativistic effects, a level with given values of L and S is split into a number of levels with different values of J. This splitting is called the *fine structure* (or the *multiplet splitting*) of the level. As we know, J takes values from $L+S$ to $|L-S|$; hence a level with given L and S is split into $2S+1$ (if $L > S$) or $2L+1$ (if $L < S$) distinct levels. Each of these is still degenerate with respect to the directions of the vector **J**; the degree of this degeneracy is $2J+1$.[†]

There is a generally accepted notation to denote the atomic energy levels (or, as they are called, the *spectral terms* of the atoms), similar to that used for the states of individual particles with definite values of the angular momentum (§29): states with different values of the

---

[†] The fine structure of energy levels in the hydrogen atom has certain special features (see §94).

total orbital angular momentum L are denoted by capital Latin letters, as follows:

$$\begin{array}{lcccccc} L = 0 & 1 & 2 & 3 & 4 & 5 & \dots \\ \quad\; S & P & D & F & G & H & \dots \end{array}$$

Above and to the left of this letter is placed the number $2S+1$, called the *multiplicity* of the term (though it must be borne in mind that this number gives the number of fine-structure components of the level only when $L \geqslant S$).[†] Below and to the right of the letter is placed the value of the total angular momentum J. Thus the symbols $^2P_{1/2}$, $^2P_{3/2}$ denote levels with $L = 1$, $S = \frac{1}{2}$, $J = \frac{1}{2}$ and $\frac{3}{2}$.

§50. Electron states in the atom

An atom with more than one electron is a complex system of mutually interacting electrons moving in the field of the nucleus. For such a system we can, strictly speaking, consider only states of the system as a whole. Nevertheless, it is found that we can, with fair accuracy, introduce the idea of the states of each individual electron in the atom, as being the stationary states of the motion of each electron in some effective centrally symmetric field due to the nucleus and to all the other electrons. These fields are in general different for different electrons in the atom, and they must all be defined simultaneously, since each of them depends on the states of all the other electrons. Such a field is said to be *self-consistent*.

Since the self-consistent field is centrally symmetric, each state of the electron is characterised by a definite value of its orbital angular momentum l. The states of an individual electron with a given l are numbered (in order of increasing energy) by the *principal quantum number n*, which takes the values $n = l+1, l+2, \dots$; this choice of the order of numbering is made in accordance with what is usual for the hydrogen atom. However, it must be noticed that the sequence of levels of increasing energy for various l in complex atoms is in

---

[†] The levels with $2S+1 = 1, 2, 3, \dots$ are called *singlet, doublet, triplet*, etc. levels.

general different from that found in the hydrogen atom. In the latter, the energy is independent of l, so that the states with larger values of n always have higher energies. In complex atoms, on the other hand, the level with $n = 5$, $l = 0$, for example, is found to lie below that with $n = 4$, $l = 2$ (this is discussed in more detail in §52).

The states of individual electrons with different values of n and l are customarily denoted by a figure which gives the value of the principal quantum number, followed by a letter which gives the value of l:[†] thus $4d$ denotes the state with $n = 4$, $l = 2$. A complete description of the atom demands that, besides the values of the total L, S and J, the states of all the electrons should also be enumerated. Thus the symbol $1s\,2p\,{}^3P_0$ denotes a state of the helium atom in which $L = 1$, $S = 1$, $J = 0$ and the two electrons are in the $1s$ and $2p$ states. If several electrons are in states with the same l and n, this is usually shown for brevity by means of an index: thus $3p^2$ denotes two electrons in the $3p$ state. The distribution of the electrons in the atom among states with different l and n is called the *electron configuration.*

For given values of n and l, the electron can have different values of the projections of the orbital angular momentum (m) and of the spin (σ). For a given l, the number m takes $2l+1$ values; the number σ is restricted to only two values, $\pm\frac{1}{2}$. Hence there are altogether $2(2l+1)$ different states with the same n and l; these states are said to be *equivalent*. According to Pauli's principle there can be only one electron in each such state. Thus at most $2(2l+1)$ electrons in an atom can simultaneously have the same n and l. An assembly of electrons occupying all the states with the given n and l is called a *closed shell* of the type concerned.

The difference in energy between atomic levels having different L and S but the same electron configuration is due to the electrostatic interaction of the electrons (we here ignore the fine structure of each multiplet level). These energy differences are usually small, and several times less than the distances between the levels of different configurations. The following empirical principle (*Hund's rule*) is known con-

---

[†] Another terminology often used is that in which electrons with principal quantum numbers $n = 1, 2, 3, \ldots$ are said to belong to the $K, L, M, \ldots$ shells.

cerning the relative position of levels with the same configuration but different L and S:

The term with the greatest possible value of S (for the given electron configuration) and the greatest possible value of L (for this S) has the lowest energy.

We shall show how the possible atomic terms can be found for a given electron configuration. If the electrons are not equivalent, the possible values of L and S are determined immediately from the rule for the addition of angular momenta. Thus, for instance, with the configurations np, $n'p$ (n, n' being different) the total angular momentum L can take the values 2, 1, 0, and the total spin $S = 0, 1$; combining these, we obtain the terms ${}^{1,3}S$, ${}^{1,3}P$, ${}^{1,3}D$.

For equivalent electrons, however, restrictions imposed by Pauli's principle considerably reduce the number of possible terms. Let us consider, for example, a configuration np^2. For $l = 1$ (the p state), the projection m of the electron orbital angular momentum can take the values $m = 1, 0, -1$, so that there are six possible states, with the following values of m and σ:

$$\begin{array}{lll} (a)\ 1, \tfrac{1}{2} & (b)\ 0, \tfrac{1}{2} & (c)\ -1, \tfrac{1}{2} \\ (a')\ 1, -\tfrac{1}{2} & (b')\ 0, -\tfrac{1}{2} & (c')\ -1, -\tfrac{1}{2}. \end{array}$$

Two electrons can be one in each of any two of these states. As a result we obtain states of the atom with the following values of the projections $M_L = \Sigma m$, $M_S = \Sigma\sigma$ of the total orbital angular momentum and spin:

$$\begin{array}{lll} (a+a')\ 2, 0 & (a+b)\ \ 1,\ 1 & (a+c)\ \ 0,\ 1 \\ & (a+b')\ \ 1,\ 0 & (a+c')\ 0,\ 0 \\ & (a'+b)\ \ 1,\ 0 & (a'+c)\ 0,\ 0 \\ & & (b+b')\ 0,\ 0. \end{array}$$

The states with M_L or M_S negative need not be written out, since they give nothing different. The presence of a state with $M_L = 2$, $M_S = 0$ shows that there must be a ${}^1D$ term, and to this term there must correspond one state (1, 0) and one (0, 0). Next, there remains a state with (1, 1), so that there must be a ^3P term; states (0, 1), (1, 0) and (0, 0) correspond to this. Finally, there remains one state (0, 0) corresponding to a ^1S term. Thus, for a configuration of two equivalent p electrons, the only possibilities are one term of each of the types ^1S, ^3P, ^1D.

For the configuration with the greatest possible number of equivalent electrons (i.e. for a closed shell), the electron angular momenta compensate each other, and the state ^1S is always the only one possible. Like terms always correspond to configurations which differ in that one of them has as many electrons as the other lacks to form a closed shell (for example, the configuration np^4 has terms of the same types as those found above for np^2). This is an evident result of the fact that the absence of an electron from the shell can be regarded as a "hole", whose state is defined by the same quantum numbers as the state of the missing electron.

§51. Fine structure of atomic levels

As has already been mentioned, the dependence of the Hamiltonian of an atom on the electron spin operators appears only when relativistic effects are taken into account, i.e. effects that become zero in the limit $c \to \infty$. We shall return in §94 to the origin of the relativistic terms in the Hamiltonian, and for the present describe the general form of these terms by means of their effects.

It is found that the relativistic terms in the Hamiltonian of an atom fall into two classes. One of these contains terms linear with respect to the spin operators of the electrons, while the other includes quadratic terms. The former correspond to the interaction between the intrinsic magnetic moments of the electrons and the magnetic moments of their orbital motion (this interaction is called *spin–orbit* interaction), while the latter correspond to the interaction between the magnetic moments of the electrons (*spin–spin* interaction). Both interactions are of the same order (the second) with respect to v/c, the ratio of the velocity of the electrons to that of light; in practice, the spin–orbit interaction considerably exceeds the spin–spin interaction in heavy atoms. This is because the spin–orbit interaction increases rapidly with the atomic number, whereas the spin–spin interaction is essentially independent of Z. This is evident from the nature of the spin–spin interaction, since this is a direct interaction between electrons and is unaffected by the field of the nucleus.

The spin–orbit interaction operator is of the form

$$\hat{V}_{sl} = \Sigma \alpha_a \hat{\mathbf{l}}_a . \hat{\mathbf{s}}_a \tag{51.1}$$

(the summation being over all the electrons in the atom), where $\hat{\mathbf{s}}_a$ are the spin operators of the electrons, $\hat{\mathbf{l}}_a$ are the operators of the orbital angular momentum of the electrons, and α_a are functions of the coordinates.

The calculation of the fine-structure energy of the atomic levels consists in averaging the perturbation operator $\hat{V}_{sl}$ over the unperturbed states of the electron shell. This averaging is done in two stages. First of all, we average over electron states of the atom with given absolute values L and S of the total orbital angular momentum and spin, but not over the directions of these. After this averaging $\hat{V}_{sl}$ is still an operator, but it must be expressible in terms of the operators of quantities that characterise the atom as a whole, not the individual electrons. Such operators are $\hat{\mathbf{S}}$ and $\hat{\mathbf{L}}$.[†]

Let the operator of the spin–orbit interaction, thus averaged, be denoted by $\hat{V}_{LS}$. Since this is linear in $\hat{\mathbf{S}}$, it has the form

$$\hat{V}_{LS} = A \hat{\mathbf{L}} . \hat{\mathbf{S}}, \tag{51.2}$$

[†] In order to clarify the meaning of this operation, it may be noted that averaging in quantum mechanics has the general significance of taking the appropriate diagonal matrix element. A partial averaging consists in taking matrix elements that are diagonal with respect to only some of the quantum numbers describing the state of the system. For example, in this case the averaging of the operator (51.1) denotes the construction of a matrix with elements $\langle nM'_L M'_S | V_{sl} | nM_L M_S \rangle$ with all possible M_L, M'_L and M_S, M'_S and diagonal with respect to all the other quantum numbers (the assembly of which we denote by n). Correspondingly, the operators $\hat{\mathbf{S}}$ and $\hat{\mathbf{L}}$ are to be regarded as matrices $\langle M'_S | \mathbf{S} | M_S \rangle$ and $\langle M'_L | \mathbf{L} | M_L \rangle$, whose elements are given by (15.11). A similar device of stepwise averaging will be needed in several subsequent treatments.

where A is a constant characterising a given (unsplit) term, i.e. depending on S and L but not on the total angular momentum J of the atom.

To calculate the energy of the splitting we must solve the secular equation formed from the matrix elements of the operator (51.2). In this case, however, we already know the correct functions in the zeroth approximation, in which the matrix of V_{LS} is diagonal. These are the wave functions of states with definite values of the total angular momentum J. The averaging with respect to such a state involves replacing the operator $\hat{\mathbf{L}}.\hat{\mathbf{S}}$ by its eigenvalue, which, according to (17.3), is

$$\mathbf{L}.\mathbf{S} = \tfrac{1}{2}[J(J+1)-L(L+1)-S(S+1)].$$

Since the values of L and S are the same for all the components of a multiplet, and we are interested only in their relative position, we can write the energy of the multiplet splitting in the form

$$\tfrac{1}{2}AJ(J+1). \tag{51.3}$$

The intervals between adjacent components (with numbers J and $J-1$) are consequently

$$\Delta E_{J,\,J-1} = AJ. \tag{51.4}$$

This formula gives what is called *Landé's interval rule.*

The constant A can be either positive or negative. For $A > 0$ the lowest component of the multiplet level is the one with the smallest possible J, i.e. $J = |L-S|$; such multiplets are said to be *normal.* If $A < 0$, on the other hand, the lowest level of the multiplet is that with $J = L+S$; these multiplets are said to be *inverted.*

For the averaged spin–spin interaction operator we should obtain, analogously to formula (51.2), an expression quadratic in $\hat{\mathbf{S}}$. The expressions $\hat{\mathbf{S}}^2$ and $(\hat{\mathbf{S}}.\hat{\mathbf{L}})^2$ are quadratic in $\hat{\mathbf{S}}$. The former has eigenvalues independent of J, and therefore does not give any splitting of the term. Hence it can be omitted, and we can write

$$\hat{V}_{SS} = B(\hat{\mathbf{S}}.\hat{\mathbf{L}})^2, \tag{51.5}$$

where B is a constant. The eigenvalues of this operator contain terms independent of J, terms proportional to $J(J+1)$, and finally a term proportional to $J^2(J+1)^2$. The first of these do not give any splitting and hence are without interest; the second can be included in the expression (51.3), which simply means a change in the constant A. Finally, the last term gives an energy

$$\tfrac{1}{4}BJ^2(J+1)^2. \tag{51.6}$$

The scheme for the construction of the atomic levels given above is based on the supposition that the orbital angular momenta of the electrons combine to give the total orbital angular momentum L of the atom, and their spins to give the total spin S. As has already been mentioned, this supposition is legitimate only when the relativistic effects are small; more exactly, the intervals in the fine structure must be small compared with the differences between levels with different L and S. This approximation is called the *Russell–Saunders case*, and we speak also of *LS coupling*.

In practice, however, this approximation has a limited range of applicability. The levels of the light atoms are arranged in accordance with the *LS* model, but as the atomic number increases the relativistic interactions in the atom become stronger, and the Russell–Saunders approximation becomes inapplicable.

In the opposite limiting case the relativistic interaction is large compared with the electrostatic. In this case we cannot speak of the orbital angular momentum and spin separately, since they are not conserved. The individual electrons are characterised by their total angular momenta j, which combine to give the total angular momentum J of the atom. This scheme of arrangement of the atomic levels is called *jj coupling*. In practice, this coupling is not found in the pure state, but various types of coupling intermediate between *LS* and *jj* are observed among the levels of very heavy atoms.

A further splitting of atomic energy levels (beyond the fine structure) results from the interaction of the magnetic moments of the electron and the nucleus, and is called the *hyperfine structure*. Since the magnetic moments of the nuclei are small in comparison with those of the electrons, this interaction is comparatively very weak, so that the intervals in the resultant splitting are very small in comparison with those in the fine structure. Hence the hyperfine structure must be considered separately for each component of the fine structure.

Let the spin of the nucleus be denoted by i (in accordance with the notation usual in atomic spectroscopy). The total angular momentum of the atom (including the nucleus) is $\mathbf{F} = \mathbf{J}+\mathbf{i}$, where $\mathbf{J}$ denotes as before the total angular momentum of the electron envelope. Each component of the hyperfine structure is described by a definite value of F. According to the general rules for the addition of angular momenta, the quantum number F takes the values

$$F = J+i, \quad J+i-1, \ldots, |J-i|. \tag{51.7}$$

§52. The Mendeleev periodic system

The elucidation of the nature of the periodic variation of properties, first observed by D. I. Mendeleev, in the series of elements when they are placed in order of increasing atomic number, requires an examination of the peculiarities in the successive completion of the electron shells of atoms. This was first carried out by N. Bohr (1922).

When we pass from one atom to the next, the charge is increased by unity and one electron is added to the envelope. At first sight we might expect the binding energy of each of the successively added electrons to vary monotonically as the atomic number increases. The actual variation, however, is entirely different.

In the normal state of the hydrogen atom there is only one electron, in the 1*s* state. In the atom of the next element, helium, another 1*s* electron is added; the binding energy of each of the 1*s* electrons in the helium atom is, however, considerably greater than in the hydrogen atom. This is a natural consequence of the difference between the field in which the electron moves in the hydrogen atom and the field encountered by an electron added to the He^+ ion. At large distances these fields are approximately the same, but near the nucleus with charge $Z = 2$ the field of the He^+ ion is stronger than that of the hydrogen nucleus with $Z = 1$. In the lithium atom ($Z = 3$), the third electron enters the 2*s* state, since no more than two electrons can be in 1*s* states at the same time. For a given *Z* the 2*s* level lies above the 1*s* level; as the nuclear charge increases, both levels become lower. In the transition from $Z = 2$ to $Z = 3$, however, the former effect is predominant, and so the binding energy of the third electron in the lithium atom is considerably less than those of the electrons in the helium atom. Next, in the atoms from Be ($Z = 4$) to Ne ($Z = 10$), first one more 2*s* electron and then six 2*p* electrons are successively added. The binding energies of these electrons increase on the average, owing to the increasing nuclear charge. The next electron added, on going to the sodium atom ($Z = 11$), enters the 3*s* state, and the binding energy again diminishes markedly, since the effect of going to a higher shell predominates over that of the increase of the nuclear charge.

This picture of the filling up of the electron envelope is characteristic of the whole sequence of elements. All the electron states can be divided into successively occupied groups such that, as the states of each group are occupied in a series of elements, the binding energy increases on the average, but when the states of the next group begin to be occupied the binding energy decreases noticeably. Fig. 11 shows

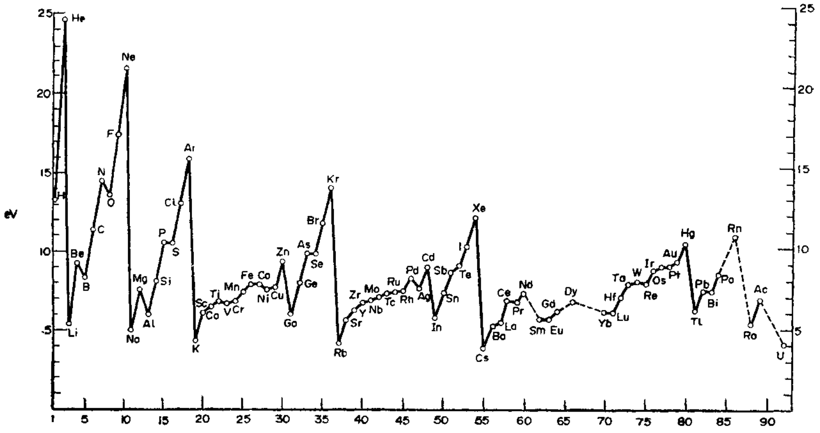

FIG. 11

those ionisation potentials of elements that are known from spectroscopic data; they give the binding energies of the electrons added as we pass from each element to the next.

The different states are distributed as follows into successively occupied groups:

$$\left.\begin{array}{ll} 1s & 2\ \text{electrons}\\ 2s,\ 2p & 8\ \text{electrons}\\ 3s,\ 3p & 8\ \text{electrons}\\ 4s,\ 3d,\ 4p & 18\ \text{electrons}\\ 5s,\ 4d,\ 5p & 18\ \text{electrons}\\ 6s,\ 4f,\ 5d,\ 6p & 32\ \text{electrons}\\ 7s,\ 6d,\ 5f,\ \ldots & \end{array}\right\} \qquad (52.1)$$

The first group is occupied in H and He; the occupation of the second and third groups corresponds to the first two (short) periods of the periodic system, containing 8 elements each. Next follow two long periods of 18 elements each, and a long period containing the rare-earth elements and 32 elements in all. The final group of states is not completely occupied in the natural (and artificial transuranic) elements.

To understand the variation of the properties of the elements as the states of each group are occupied, the following property of *d* and *f* states, which distinguishes them from *s* and *p* states, is important. The curves of the effective potential energy of the centrally symmetric field (composed of the electric field and the centrifugal field) for and electron in a heavy atom have a rapid and almost vertical drop to a deep minimum near the origin; they then begin to rise, and approach zero asymptotically. For *s* and *p* states, the rising parts of these curves are very close together. This means that the electron is at approximately the same distance from the nucleus in these states. The curves for the *d* states, and particularly for the *f* states, on the other hand, pass considerably further to the left; the "classically accessible" region which they delimit ends considerably closer in than that for the *s* and *p* states with the same total electron energy. In other words, an electron in the d and f states is mainly much closer to the nucleus than in the s and p states.

Many properties of atoms (including the chemical properties of elements; see §58) depend principally on the outer regions of the electron envelopes. The above characteristic of the d and f states is very important in this connection. Thus, for instance, when the $4f$ states are being filled (in the rare-earth elements; see below), the added electrons are located considerably closer to the nucleus than those in the states previously occupied. As a result, these electrons have practically no effect on the chemical properties, and all the rare-earth elements are chemically very similar.

The elements containing complete d and f shells (or not containing these shells at all) are called elements of the *principal groups*; those in which the filling up of these states is actually in progress are called elements of the *intermediate groups*. These groups of elements are conveniently considered separately.

Let us begin with the elements of the principal groups. Hydrogen and helium have the following normal states:

$$_1\mathrm{H} : 1s\,^2S_{1/2} \qquad _2\mathrm{He} : 1s^2\,^1S_0$$

(the number with the chemical symbol always signifies the atomic number). The electron configurations of the remaining elements of the principal groups are shown in Table 1.

In each atom, the shells shown on the right of the table in the same line and above are completely filled. The electron configuration in the shells that are being filled is shown at the top, while the principal quantum number of the electrons in these states is shown by the figure on the left of the table in the same line. The normal states of the whole atom are shown at the bottom. Thus, the aluminium atom has the electron configuration $1s^2\,2s^2\,2p^6\,3s^2\,3p\,^2P_{1/2}$.

The atoms of the inert gases (He, Ne, Ar, Kr, Xe, Rn) occupy a special position in the table: the filling up of one of the groups of states listed in (52.1) is completed in each of them. Their electron configurations have unusual stability (their ionisation potentials are

TABLE 1

Electron configurations of the atoms of elements in the principal groups

	s	s^2	s^2p	s^2p^2	s^2p^3	s^2p^4	s^2p^5	s^2p^6	
$n = 2$	$_{3}$Li	$_{4}$Be	$_{5}$B	$_{6}$C	$_{7}$N	$_{8}$O	$_{9}$F	$_{10}$Ne	$1s^2$
3	$_{11}$Na	$_{12}$Mg	$_{13}$Al	$_{14}$Si	$_{15}$P	$_{16}$S	$_{17}$Cl	$_{18}$Ar	$2s^2\,2p^6$
4	$_{19}$K	$_{20}$Ca							$3s^2\,3p^6$
4	$_{29}$Cu	$_{30}$Zn	$_{31}$Ga	$_{32}$Ge	$_{33}$As	$_{34}$Se	$_{35}$Br	$_{36}$Kr	$3d^{10}$
5	$_{37}$Rb	$_{38}$Sr							$4s^2\,4p^6$
5	$_{47}$Ag	$_{48}$Cd	$_{49}$In	$_{50}$Sn	$_{51}$Sb	$_{52}$Te	$_{53}$I	$_{54}$Xe	$4d^{10}$
6	$_{55}$Cs	$_{56}$Ba							$5s^2\,5p^6$
6	$_{79}$Au	$_{80}$Hg	$_{81}$Tl	$_{82}$Pb	$_{83}$Bi	$_{84}$Po	$_{85}$At	$_{86}$Rn	$4f^{14}\,5d^{10}$
7	$_{87}$Fr	$_{88}$Ra							$6s^2\,6p^6$
	$^2S_{1/2}$	$^1S_0$	$^2P_{1/2}$	$^3P_0$	$^4S_{3/2}$	$^3P_2$	$^2P_{3/2}$	$^1S_0$	

the greatest in their respective series). This causes the chemical inertness of these elements.

We see that the occupation of different states occurs very regularly in the series of elements of the principal groups: first the s states and then the p states are occupied for each principal quantum number n. The electron configurations of the ions of these elements are also regular (until electrons from the d and f shells are removed in the ionisation): each ion has the configuration corresponding to the preceding atom. Thus, the Mg^+ ion has the configuration of the sodium atom, and the Mg^{++} ion that of neon.

Let us now turn to the elements of the intermediate groups. The filling up of the $3d$, $4d$, and $5d$ shells takes place in groups of elements called respectively the *iron group*, the *palladium group* and the *platinum group*. Table 2 gives those electron configurations and terms of the atoms in these groups that are known from experimental spectroscopic data. As is seen from this table, the d shells are filled up with considerably less regularity than the s and p shells in the atoms of elements of the principal groups. Here a characteristic feature is the "competi-

TABLE 2

Electron configurations of the atoms of elements in the iron, palladium and platinum groups

Iron group

	$_{21}$Sc	$_{22}$Ti	$_{23}$V	$_{24}$Cr	$_{25}$Mn	$_{26}$Fe	$_{27}$Co	$_{28}$Ni
Ar envelope +	$3d\,4s^2$ $^2D_{3/2}$	$3d^2\,4s^2$ $^3F_2$	$3d^3\,4s^2$ $^4F_{3/2}$	$3d^5\,4s$ $^7S_3$	$3d^5\,4s^2$ $^6S_{5/2}$	$3d^6\,4s^2$ $^5D_4$	$3d^7\,4s^2$ $^4F_{9/2}$	$3d^8\,4s^2$ $^3F_4$

Palladium group

	$_{39}$Y	$_{40}$Zr	$_{41}$Nb	$_{42}$Mo	$_{43}$Tc	$_{44}$Ru	$_{45}$Rh	$_{46}$Pd
Kr envelope +	$4d\,5s^2$ $^2D_{3/2}$	$4d^2\,5s^2$ $^3F_2$	$4d^4\,5s$ $^6D_{1/2}$	$4d^5\,5s$ $^7S_3$	$4d^5\,5s^2$ $^6S_{5/2}$	$4d^7\,5s$ $^5F_5$	$4d^8\,5s$ $^4F_{9/2}$	$4d^{10}$ $^1S_0$

Platinum group

	$_{57}$La
Xe envelope +	$5d\,6s^2$ $^2D_{3/2}$

	$_{71}$Lu	$_{72}$Hf	$_{73}$Ta	$_{74}$W	$_{75}$Re	$_{76}$Os	$_{77}$Ir	$_{78}$Pt
Xe envelope $+4f^{14}+$	$5d\,6s^2$ $^2D_{3/2}$	$5d^2\,6s^2$ $^3F_2$	$5d^3\,6s^2$ $^4F_{3/2}$	$5d^4\,6s^2$ $^5D_0$	$5d^5\,6s^2$ $^6S_{5/2}$	$5d^6\,6s^2$ $^5D_4$	$5d^7\,6s^2$ $^4F_{9/2}$	$5d^9\,6s$ $^3D_3$

tion" between the *s* and *d* states. It is seen in the fact that, instead of a regular sequence of configurations of the type $d^p s^2$ with increasing p, configurations of the type $d^{p+1}s$ or d^{p+2} are often found. Thus, in the iron group, the chromium atom has the configuration $3d^5\,4s$, and not $3d^4\,4s^2$; after nickel with 8 *d* electrons, there follows at once the copper atom with a completely filled *d* shell (and hence we place this element in the principal groups). This lack of regularity is observed in the terms of ions also: the electron configurations of the ions do not usually agree with those of the preceding atoms. For instance, the V^+ ion has the configuration $3d^4$ (and not $3d^2\,4s^2$ like titanium); the Fe^+ ion has $3d^6\,4s$ (instead of $3d^5\,4s^2$ as in manganese). We may remark that all ions found naturally in crystals and solutions contain only *d* (not *s* or *p*) electrons in their incomplete shells. Thus iron is found in crystals or solutions only as the ions Fe^{++} and Fe^{+++}, whose configurations are $3d^6$ and $3d^5$ respectively.

A similar situation occurs in the filling up of the $4f$ shell; this takes place in the series of elements known as the *rare earths* (Table 3). The filling up of the $4f$ shell also occurs in a slightly irregular manner characterised by the "competition" between $4f$, $5d$ and $6s$ states.

The last group of intermediate elements begins with actinium. In this group the $6d$ and $5f$ shells are filled, similarly to what happens in the group of rare-earth elements.

§53. **X-ray terms**

The binding energy of the inner electrons in the atom is so large that, if such an electron makes a transition into an outer unfilled shell (or is removed from the atom), the excited atom (or ion) is mechanically unstable with respect to ionisation, which is accompanied by the reconstruction of the electron envelope and the formation of a stable ion. However, because of the comparatively weak interaction between the electrons in the atom, the probability of such a transition is comparatively small, so that the lifetime τ of the excited state is long. Hence the width $\hbar/\tau$ of the level (see §38) is so small that it is reasonable to regard the energies of an atom with an excited inner electron as discrete

TABLE 3

Electron configurations of the atoms of the rare-earth elements

	$_{58}$Ce	$_{59}$Pr	$_{60}$Nd	$_{61}$Pm	$_{62}$Sm	$_{63}$Eu	$_{64}$Gd	$_{65}$Tb	$_{66}$Dy	$_{67}$Ho	$_{68}$Er	$_{69}$Tm	$_{70}$Yb
Xe envelope +	$4f5d\,6s^2$	$4f^3\,6s^2$	$4f^4\,6s^2$	$4f^5\,6s^2$	$4f^6\,6s^2$	$4f^7\,6s^2$	$4f^7\,5d\,6s^2$	$4f^9\,6s^2$	$4f^{10}\,6s^2$	$4f^{11}\,6s^2$	$4f^{12}\,6s^2$	$4f^{13}\,6s^2$	$4f^{14}\,6s^2$
	$^1G_4$	$^4I_{9/2}$	$^5I_4$	$^6H_{5/2}$	$^7F_0$	$^8S_{7/2}$	$^9D_2$	$^6H_{15/2}$	$^5I_8$	$^4I_{15/2}$	$^3H_6$	$^2F_{7/2}$	$^1S_0$

energy levels of "quasi-stationary" states of the atom. These levels are called *X-ray terms.*[†]

The X-ray terms are primarily classified according to the shell from which the electron is removed, or in which, as we say, a "hole" is formed. Where the electron goes has almost no effect on the energy of the atom, and hence is unimportant.

The total angular momentum of the set of electrons occupying any shell is zero. When one electron has been removed, the shell acquires some angular momentum j. For the (n, l) shell, the angular momentum j can take the values $l \pm \frac{1}{2}$. Thus we obtain levels which might be denoted by $1s_{1/2}$, $2s_{1/2}$, $2p_{1/2}$, $2p_{3/2}$, ..., where the value of j is added as a suffix to the letter giving the position of the "hole". It is usual, however, to employ special symbols as follows:

$1s_{1/2}$	$2s_{1/2}$	$2p_{1/2}$	$2p_{3/2}$	$3s_{1/2}$	$3p_{1/2}$	$3p_{3/2}$	$3d_{3/2}$	$3d_{5/2}$ ...
K	$L_{\rm I}$	$L_{\rm II}$	$L_{\rm III}$	$M_{\rm I}$	$M_{\rm II}$	$M_{\rm III}$	$M_{\rm IV}$	$M_{\rm V}$ ...

Levels with the same n (denoted by the same capital letter) lie close together and at a distance from levels with a different n. The reason for this is that, owing to the relative nearness of the inner electrons to the nucleus, they are in the almost unscreened Coulomb field of the nucleus, and hence their states are "hydrogen-like"; the energy is approximately that of a single electron in the field of a nucleus with charge Ze, and so depends only on the principal quantum number n (§31). If relativistic effects are taken into account, terms with different j are separated, such as, for example, $L_{\rm I}$ and $L_{\rm II}$ from $L_{\rm III}$, and $M_{\rm I}$ and $M_{\rm II}$ from $M_{\rm III}$ and $M_{\rm IV}$. These pairs of levels are said to be *relativistic* doublets. The separation of terms with different l and the same j (for instance $L_{\rm I}$ and $L_{\rm II}$, $M_{\rm I}$ and $M_{\rm II}$) is due to the deviation of the field in which the inner electrons move from the Coulomb field of the nucleus, i.e. to the taking into account of the interaction of the electron with other electrons. These are said to be *screening* doublets.

The width of an X-ray term is determined by the total probability of all possible processes of rearrangement of the electron envelope

---

[†] The name is due to the fact that transitions between these levels cause the emission of X-rays by the atom.

of the atom so as to fill the "hole" in question. In the heavy atoms, transitions of the hole from a given shell to a higher one (i.e. electron transitions in the opposite direction) are the most important, and are accompanied by the emission of X-ray quanta. The probability of these "radiative" transitions, and therefore the corresponding part of the level width, increase very rapidly with the atomic number.

For lighter atoms, an important or even predominant part in determining the level width is played by radiationless transitions in which the energy liberated when a hole is filled by an electron from above goes to remove another inner electron from the atom (called the *Auger effect*). As a result of this process the atom is in a state with two holes.

§54. An atom in an electric field

In classical theory, the electrical properties of a system of particles are described by its electric multipole moments of various orders (see *Mechanics and Electrodynamics*, §§62, 63). In the quantum theory, the definitions of these quantities are the same in form, but they must now be regarded as operators.

The first multipole moment is the *dipole moment*, defined as the vector

$$\mathbf{d} = \Sigma e\mathbf{r}. \tag{54.1}$$

For an atom (whose nucleus is assumed to be fixed at the origin), the summation is over all the electrons; the suffix which numbers the electrons is omitted for brevity. The mean value of the dipole moment in a stationary state of the atom is obtained by averaging the operator (54.1) over the wave function of the state, i.e. by taking the corresponding diagonal matrix element. The matrix of this operator, like that of any polar vector (see §19), has non-zero elements only for transitions between states of different parity. The diagonal elements are therefore always zero. In other words, the mean values of the dipole moment of an atom in stationary states are zero.†

---

† Here it is assumed that the energy levels of the atom are degenerate only with respect to the directions of its total angular momentum. All states which differ

The *quadrupole moment* of a system is defined as the symmetrical tensor

$$Q_{ik} = \sum e(3x_i x_k - \delta_{ik}\mathbf{r}^2), \tag{54.2}$$

the sum of whose diagonal terms is zero.

It must be noted first of all that the mean value of the quadrupole moment is zero in any state having total angular momentum $J = 0$ or $\frac{1}{2}$. This can be proved by the method described in §18 for finding the selection rules for the matrix elements of vectors and tensors. Using this method, we formally assign to the tensor (54.2) an "angular momentum" $L = 2$. The matrix element is non-zero if the addition of this to the angular momenta J_1 and J_2 of the initial and final states can give the result zero. This cannot occur with the three values 2, 0, 0 or 2, $\frac{1}{2}$, $\frac{1}{2}$, and the diagonal matrix elements therefore vanish for $J_1 = J_2 = 0$ or $\frac{1}{2}$.

For a state of the atom having a given total angular momentum J, the mean values of the quadrupole moment depend on the angular momentum component M_J. The dependence can be found as follows.

The averaging of the operator (54.2) with respect to the state of the atom is conveniently carried out in two stages (cf. §51). We first average over states for which the value of J is fixed, but not that of M_J. The operator thus averaged, denoted by $\hat{Q}_{ik}$, must be expressible in terms of operators of quantities describing the state of the atom as a whole. The only such vector is the "vector" $\hat{\mathbf{J}}$. Thus $\hat{Q}_{ik}$ must have the form

$$\hat{Q}_{ik} = \frac{3Q}{2J(2J-1)}(\hat{J}_i\hat{J}_k + \hat{J}_k\hat{J}_i - \tfrac{2}{3}\delta_{ik}\hat{\mathbf{J}}^2), \tag{54.3}$$

where the expression in parentheses is constructed so as to be symmetrical in the suffixes i and k and to vanish on summation over $i = k$;

---

only in the values of the projection of the total angular momentum have the same parity, and therefore any superposition of them has this parity also. The hydrogen atom forms an exceptional case here, since its levels have an additional "accidental" degeneracy. The mutually degenerate states with different values of the orbital angular momentum l may have different parities. From their wave functions one can obtain superpositions having no definite parity; the corresponding diagonal matrix element of the dipole moment need not be zero.

the significance of the coefficient Q will be explained later. The operators $\hat{J}_i$ must here be understood as the familiar (§15) matrices with respect to states having different values of M_J.

Since the three components of the angular momentum vector cannot simultaneously have definite values, the same is true of the components of the tensor (54.3). For the component $\hat{Q}_{zz}$, we have

$$\hat{Q}_{zz} = \frac{3Q}{J(2J-1)}(\hat{J}_z^2 - \tfrac{1}{3}\hat{\mathbf{J}}^2).$$

The averaging of this operator with respect to a state having fixed values of J and M_J now simply means replacing the operators by their eigenvalues. Thus we find

$$\overline{Q_{zz}} = \frac{3Q}{J(2J-1)}[M_J^2 - \tfrac{1}{3}J(J+1)], \tag{54.4}$$

and this gives the required dependence. For $M_J = J$ (when the angular momentum is "entirely" in the z-direction), we have $\bar{Q}_{zz} = Q$; this quantity is usually called simply the *quadrupole moment*.

Let us consider an atom in a uniform electric field $\mathbf{E}$. In such an atom, the electrons are in an axially symmetric field (the field of the nucleus together with the uniform field). The total angular momentum $\mathbf{J}$ of the atom is therefore no longer conserved, but its projection on the direction of the axis of symmetry (taken as the z-axis) is conserved.

By distinguishing a particular direction in space, the external field removes the degeneracy of the levels with respect to the directions of the angular momentum: the states having different values of $J_z = M_J$, whose energies are equal in the free atom, have different energies in the electric field (the *Stark effect*). The splitting of the levels is, however, incomplete; the energies of the states which differ only in the sign of M_J remain the same. This is a direct consequence of the symmetry under time reversal (§23). Since the directions of all velocities are reversed by this operation, so is the sign of the angular-momentum projection, the energy of the system remaining unchanged; the field $\mathbf{E}$ is unaltered (see *Mechanics and Electrodynamics*, §44).

Thus the energy levels of an atom in an electric field remain doubly degenerate, except for the levels having $M_J = 0$. But if the total angular momentum J is half-integral, the value $M_J = 0$ cannot occur, and all the levels are then doubly degenerate. This is a particular case of a more general rule: it can be shown from the requirements of symmetry under time reversal that, for a system with half-integral J, the double degeneracy of the levels is retained in any electric field (not necessarily uniform). This is *Kramers' theorem.*†

If the electric field is so weak that the additional energy due to it is small compared with the distances between neighbouring unperturbed energy levels of the atom, then perturbation theory can be used to calculate the displacement of the levels. In a uniform field, the perturbation operator is the potential energy of the atom in the field, expressed in terms of its dipole moment:

$$V = -\mathbf{E}.\mathbf{d} = -|\mathbf{E}|\, d_z. \tag{54.5}$$

In the first approximation, the displacement of the energy levels is determined by the corresponding diagonal matrix elements of the perturbation operator. These elements are zero, however, because the mean values of the dipole moment are zero. The splitting of the levels in the electric field therefore occurs only in the second approximation of perturbation theory, and accordingly is proportional to the square of the field.‡

Being a quadratic function of the field, the displacement ΔE_n of the level E_n must be expressed by a formula of the type

$$\Delta E_n = -\tfrac{1}{2}\alpha_{ik}^{(n)} E_i E_k, \tag{54.6}$$

where the coefficients $\alpha_{ik}^{(n)}$ form a symmetrical tensor of rank two;

---

† It must be emphasised, however, that in an arbitrary electric field the states of the atom can no longer be described by the values of the angular-momentum projection, since in an inhomogeneous field the components of the angular momentum as well as its absolute value are not conserved.

‡ The hydrogen atom forms an exception; for its stationary states, the mean value of the dipole moment need not be zero. The splitting of the energy levels of the hydrogen atom is consequently linear in the field.

taking the z-axis in the direction of the field, we have

$$\Delta E_n = -\tfrac{1}{2}\alpha_{zz}^{(n)}|\mathbf{E}|^2. \tag{54.7}$$

The coefficients in these formulae have a further significance: they represent the polarisability of the atom in the external electric field. This follows from the general formula

$$(\partial\hat{H}/\partial\lambda)_{nn} = \partial E_n/\partial\lambda. \tag{54.8}$$

The expression on the left is the diagonal matrix element of the operator $\partial\hat{H}/\partial\lambda$, where $\hat{H}$ is the Hamiltonian of the system, a function of some parameter λ; the eigenvalues E_n of the Hamiltonian are also functions of λ. If the parameter λ in (54.8) is taken to be the magnitude $|\mathbf{E}|$ of the field, and if we put

$$\hat{H} = \hat{H}_0+\hat{V} = \hat{H}_0-|\mathbf{E}|\,d_z,$$

then, with (54.7),

$$\bar{d}_z = \alpha_{zz}^{(n)}|\mathbf{E}|. \tag{54.9}$$

The *polarisability* of the atom is the proportionality coefficient between its dipole moment in a field and the magnitude of the field.

To prove formula (54.8), we begin from the equation

$$(\hat{H}-E_n)\psi_n = 0$$

which determines the eigenvalues of the operator $\hat{H}$. Differentiating this with respect to λ and then multiplying on the left by ψ_n^*, we obtain

$$\psi_n^*(\hat{H}-E_n)\frac{\partial\psi_n}{\partial\lambda} = \psi_n^*\left(\frac{\partial E_n}{\partial\lambda}-\frac{\partial\hat{H}}{\partial\lambda}\right)\psi_n.$$

On integration with respect to q, the left-hand side gives zero, since the operator $\hat{H}$ is Hermitian and therefore (see (3.10))

$$\int\psi_n^*(\hat{H}-E_n)\frac{\partial\psi_n}{\partial\lambda}\,dq = \int\frac{\partial\psi_n}{\partial\lambda}(\hat{H}^*-E_n)\psi_n^*\,dq,$$

and $(\hat{H}^*-E_n)\psi_n^* = 0$. The right-hand side gives the required formula.

§55. An atom in a magnetic field

Let us consider an atom in a uniform magnetic field **H**. According to (43.4), its Hamiltonian is

$$\hat{H} = \frac{1}{2m}\sum_a \left[\hat{\mathbf{p}}_a + \frac{|e|}{c}\mathbf{A}(\mathbf{r}_a)\right]^2 + U + \frac{\hbar|e|}{mc}\mathbf{H}\,.\,\hat{\mathbf{S}}, \tag{55.1}$$

where the summation is taken over all the electrons (the electron charge being written as $e = -|e|$), U is the energy of interaction of the electrons with the nucleus and with another, and $\hat{\mathbf{S}} = \sum \hat{\mathbf{s}}_a$ is the operator of the total (electron) spin of the atom.

We can take the vector potential of the uniform field in the form

$$\mathbf{A} = \tfrac{1}{2}\mathbf{H}\times\mathbf{r} \tag{55.2}$$

(see *Mechanics and Electrodynamics*, §46). It is easy to see that, with this choice, the operator $\hat{\mathbf{p}} = -i\hbar\nabla$ commutes with **A**, since for any function $\psi(\mathbf{r})$

$$(\hat{\mathbf{p}}\mathbf{A} - \mathbf{A}\hat{\mathbf{p}}) = -i\hbar\nabla(\mathbf{A}\psi) + i\hbar\mathbf{A}\,.\,\nabla\psi = -i\hbar\psi\,\mathrm{div}\,\mathbf{A},$$

i.e.

$$\hat{\mathbf{p}}\mathbf{A} - \mathbf{A}\hat{\mathbf{p}} = -i\hbar\,\mathrm{div}\,\mathbf{A}.$$

For the vector (55.2), $\mathrm{div}\,\mathbf{A} = -\tfrac{1}{2}\mathbf{H}.\mathrm{curl}\,\mathbf{r} = 0$. Using this result in expanding the bracket in (55.1), we can rewrite the Hamiltonian in the form

$$\hat{H} = \hat{H}_0 + \frac{|e|}{mc}\sum_a \mathbf{A}_a\,.\,\hat{\mathbf{p}}_a + \frac{e^2}{2mc^2}\sum_a \mathbf{A}_a^2 + \frac{|e|\hbar}{mc}\mathbf{H}\,.\,\hat{\mathbf{S}},$$

where $\hat{H}_0$ is the Hamiltonian of the atom in the absence of the field. Substituting **A** from (55.2), we obtain

$$\hat{H} = \hat{H}_0 + \frac{|e|}{2mc}\mathbf{H}\,.\,\sum_a \mathbf{r}_a\times\hat{\mathbf{p}}_a + \frac{e^2}{8mc^2}\sum_a (\mathbf{H}\times\mathbf{r}_a)^2 + \frac{|e|\hbar}{mc}\mathbf{H}\,.\,\hat{\mathbf{S}}.$$

The vector product $\mathbf{r}_a\times\hat{\mathbf{p}}_a$, however, is the operator of the electron orbital angular momentum, and the summation over all the electrons gives the operator $\hbar\hat{\mathbf{L}}$ of the total orbital angular momentum of the atom. Thus

$$\hat{H} = \hat{H}_0 + \mu_B(\hat{\mathbf{L}}+2\hat{\mathbf{S}}).\mathbf{H} + \frac{e^2}{8mc^2}\sum_a(\mathbf{H}\times\mathbf{r}_a)^2, \qquad (55.3)$$

where μ_B is the Bohr magneton.

Like an electric field, the external magnetic field splits the atomic levels, removing the degeneracy with respect to the directions of the total angular momentum (the *Zeeman effect*). Let us determine the amount of this splitting for atomic levels having definite values of the quantum numbers J, L and S (i.e. assuming the case of LS coupling; see §51).

We shall assume that the magnetic field is so weak that $\mu_B|\mathbf{H}|$ is small compared with the distances between the energy levels of the atom, including the fine-structure intervals. Then the second and third terms in (55.3) can be regarded as a perturbation, the unperturbed levels being the separate components of the multiplets. In the first approximation we can neglect the third term, which is quadratic with respect to the field, in comparison with the second term, which is linear.

In the first approximation of perturbation theory, the energy ΔE of the splitting is determined by the mean values of the perturbation in (unperturbed) states which have different values of the projection of the total angular momentum on the direction of the field. Taking this direction as the z-axis, we have

$$\Delta E = \mu_B|\mathbf{H}|(\bar{L}_z+2\bar{S}_z) = \mu_B|\mathbf{H}|(\bar{J}_z+\bar{S}_z). \qquad (55.4)$$

The mean value $\overline{J_z}$ is just the given eigenvalue of $J_z = M_J$. The mean value $\overline{S_z}$ can be found as follows, using stepwise averaging (cf. §51).

We first average the operator $\hat{\mathbf{S}}$ over a state of the atom with fixed values of S, L and J, but not of M_J. The operator $\bar{\hat{\mathbf{S}}}$ thus averaged must be "parallel" to $\hat{\mathbf{J}}$, the only conserved "vector" characterising a free atom. We can therefore write

$$\bar{\mathbf{S}} = \text{constant}\times\mathbf{J}.$$

In this form, however, the equation is purely conventional, since the three components of the vector **J** cannot simultaneously have definite values. Its z-component can be taken literally:

$$\overline{S_z} = \text{constant}\times J_z = \text{constant}\times M_J,$$

as can the equation

$$\overline{\mathbf{S}}\,.\,\mathbf{J} = \text{constant}\times \mathbf{J}^2 = \text{constant}\times J(J+1),$$

which is obtained on multiplying both sides by **J**. Taking the conserved vector **J** under the averaging sign gives $\overline{\mathbf{S}}\,.\,\mathbf{J} = \overline{\mathbf{S}\,.\,\mathbf{J}}$. The mean value $\overline{\mathbf{S}\,.\,\mathbf{J}}$ is the same as the eigenvalue

$$\mathbf{S}\,.\,\mathbf{J} = \tfrac{1}{2}[J(J+1)-L(L+1)+S(S+1)],$$

to which it is equal in a state having definite values of $\mathbf{L}^2$, $\mathbf{S}^2$, and $\mathbf{J}^2$ (from formula (17.3), in which we must take $\mathbf{L}_1$, $\mathbf{L}_2$, and $\mathbf{L}$ as **S**, **L**, and **J** respectively). Determining the constant from the second equation and substituting in the first equation, we therefore have

$$S_z = M_J\,\mathbf{J}\,.\,\mathbf{S}/J^2. \tag{55.5}$$

Collecting the above expressions and substituting in (55.4), we find the following final expression for the energy of the splitting:

$$\Delta E = \mu_B g M_J\,|\mathbf{H}|, \tag{55.6}$$

where

$$g = 1+\frac{J(J+1)-L(L+1)+S(S+1)}{2J(J+1)} \tag{55.7}$$

is what is called the *Landé factor* or *gyromagnetic factor*. If there is no spin ($S = 0$, and so $J = L$), $g = 1$; if $L = 0$, and so $J = S$, then $g = 2$.

Formula (55.6) gives different values of the energy for all the $2J+1$ values $M_J = J, J-1, \ldots, -J$. Thus the magnetic field completely removes the degeneracy of the levels with respect to directions of the angular momentum, unlike the electric field, which leaves the levels with $M_J = \pm |M_J|$ unsplit.[†] However, the linear splitting described by (55.6) does not occur if $g = 0$; this can refer even to states for which $J \neq 0$, such as ${}^4D_{1/2}$.

We have seen in §54 that there is a relation between the displacement of an energy level of an atom in an electric field and its mean dipole moment. A similar relation exists in the magnetic case. The potential energy of a system of charges in a uniform magnetic field is given, in classical theory, by $-\boldsymbol{\mu}.\mathbf{H}$, where $\boldsymbol{\mu}$ is the magnetic moment of the system. In the quantum theory, it is replaced by the corresponding operator, so that the Hamiltonian of the system is

$$\hat{H} = \hat{H}_0 - \hat{\boldsymbol{\mu}}.\mathbf{H} = \hat{H}_0 - \hat{\mu}_z |\mathbf{H}|.$$

Now, applying (54.8), with the field $|\mathbf{H}|$ as the parameter λ, we find that the mean value of the magnetic moment is

$$\bar{\mu}_z = -\partial \Delta E / \partial |\mathbf{H}|, \tag{55.8}$$

where ΔE is the displacement of the energy level for the given state of the atom. Substituting (55.6), we see that an atom in a state with a definite value M_J of the projection of the total angular momentum on some direction z has a mean magnetic moment in that direction

$$\bar{\mu}_z = -\mu_B g M_J. \tag{55.9}$$

If the atom has neither spin nor orbital angular momentum ($S = L = 0$), the second term in (55.3) gives no displacement of the level in either the first or higher approximations (since all the matrix elements of **L** and **S** vanish). The entire effect in this case therefore arises from the third term in (55.3), and in the first approximation of perturbation theory the displacement of the level is equal to the mean value

$$\Delta E = \frac{e^2}{8mc^2} \sum_a \overline{(\mathbf{H} \times \mathbf{r}_a)^2}. \tag{55.10}$$

[†] The arguments applied to the electric-field case in this respect in §54 are not valid for a magnetic field. The reason is that the operation of time reversal has to be accompanied by the change $\mathbf{H} \rightarrow -\mathbf{H}$ (see *Mechanics and Electrodynamics*, §44). The states obtained from each other by this operation therefore belong to atoms in different fields, not in the same field.

Putting $(\mathbf{H}\times\mathbf{r}_a)^2 = \mathbf{H}^2 r_a^2 \sin^2\theta_a$, where θ_a is the angle between $\mathbf{H}$ and $\mathbf{r}_a$, we average with respect to the directions of $\mathbf{r}_a$. A state of an atom with $L = S = 0$ is spherically symmetrical; the averaging with respect to directions is therefore independent of that with respect to distances r_a, and gives $\overline{\sin^2\theta_a} = 1-\overline{\cos^2\theta_a} = \frac{2}{3}$. Thus

$$\Delta E = \frac{e^2}{12mc^2}\mathbf{H}^2\sum_a \overline{r_a^2}. \qquad (55.11)$$

The magnetic moment of the atom calculated from (55.8) is then proportional to the field (an atom with $L = S = 0$ has, of course, no magnetic moment in the absence of a field). Writing it in the form $\chi|\mathbf{H}|$, we can regard the coefficient χ as the magnetic susceptibility of the atom, given by *Langevin's formula*

$$\chi = -\frac{e^2}{6mc^2}\sum_a \overline{r_a^2}. \qquad (55.12)$$

It is negative, i.e. the atom is diamagnetic.

CHAPTER 8

THE DIATOMIC MOLECULE

§56. Electron terms in the diatomic molecule

In the theory of molecules an important part is played by the fact that the masses of atomic nuclei are very large compared with those of the electrons. Because of this difference in mass, the rates of motion of the nuclei in the molecule are small in comparison with the velocities of the electrons. This makes it possible to regard the motion of the electrons as being about fixed nuclei placed at given distances from one another. On determining the energy levels U_n for such a system, we find what are called the *electron terms* for the molecule. Unlike those for atoms, where the energy levels were certain numbers, the electron terms here are not numbers but functions of parameters, the distances between the nuclei in the molecule. The energy U_n includes also the electrostatic energy of the mutual interaction of the nuclei, so that U_n is the total energy of the molecule for a given arrangement of the fixed nuclei.

The most complete theoretical investigation is possible for the simplest type, the *diatomic* molecules, which will be discussed in this chapter. The electron terms of a diatomic molecule are functions of only one parameter, the distance r between the nuclei.

One of the chief principles in the classification of the atomic terms was the classification according to the values of the total orbital angular momentum L. In molecules, however, there is no law of conservation of the total orbital angular momentum of the electrons, since the electric field of several nuclei is not centrally symmetric.

In diatomic molecules, however, the field has axial symmetry about an axis passing through the two nuclei. Hence the projection of the orbital angular momentum on this axis is here conserved, and we can classify the electron terms of the molecules according to the values of this projection. The absolute value of the projected orbital angular momentum along the axis of the molecule is customarily denoted by the letter Λ; it takes the values 0, 1, 2, ... The terms with different values of Λ are denoted by the capital Greek letters corresponding to the Latin letters for the atomic terms with various L. Thus, for $\Lambda = 0, 1, 2$ we speak of Σ, Π and Δ terms respectively.

Next, each electron state of the molecule is characterised by the total spin S of all the electrons in the molecule. If all relativistic interactions (i.e. the fine structure of the term; cf. §51) are neglected, an electron term with spin S has degeneracy of degree $2S+1$ with respect to the directions of the total spin. The number $2S+1$ is, as in atoms, called the *multiplicity* of the term, and is written as an index before the letter for the term; thus $^3\Pi$ denotes a term with $\Lambda = 1$, $S = 1$.

Besides rotations through any angle about the axis, the symmetry of the molecule allows also a reflection in any plane passing through the axis. If we effect such a reflection, the energy of the molecule is unchanged. The state obtained from the reflection is, however, not completely identical with the initial state. For, on reflection in a plane passing through the axis of the molecule, the sign of the angular momentum about this axis is changed.† Thus we conclude that all electron terms with non-zero values of Λ are doubly degenerate: to each value of the energy, there correspond two states which differ in the direction of the projection of the orbital angular momentum on the axis of the molecule. In the case where $\Lambda = 0$ the state of the molecule is not changed at all on reflection, so that the Σ terms are not degenerate. The wave function of a Σ term can only be multiplied by a constant as a result of the reflection. Since a double reflection in the same plane is an identity transformation, this constant is ± 1.

---

† Let the reflection be in the xz-plane, the z-axis being the axis of the molecule. Under this transformation, only the y-components of the vectors $\mathbf{r}$ and $\mathbf{p}$ change sign, and $(\mathbf{r}\times\mathbf{p})_z = xp_y - yp_x$ therefore changes sign also.

Thus we must distinguish Σ terms whose wave functions are unaltered on reflection and those whose wave functions change sign. The former are denoted by Σ^+, and the latter by Σ^-.

If the molecule consists of two similar atoms, a new symmetry appears, and with it an additional characteristic of the electron terms. A diatomic molecule with identical nuclei has a centre of symmetry at the point bisecting the line joining the nuclei. (We shall take this point as the origin.) Hence the Hamiltonian is invariant with respect to a simultaneous change of sign of the coordinates of all the electrons in the molecule (the coordinates of the nuclei remaining unchanged). Since the operator of this transformation also commutes with the orbital angular momentum operator, we have the possibility of classifying terms with a given value of Λ according to their parity: the wave functions of *even* (g) states are unchanged when the coordinates of the electrons change sign, while those of *odd* (u) states change sign. The suffixes u, g indicating the parity are customarily written with the letter for the term: Π_u, Π_g, and so on.

There is an empirical rule, according to which the normal electron state in the great majority of chemically stable diatomic molecules is completely symmetrical: the electron wave function is invariant with respect to all symmetry transformations in the molecule. As we shall show in §58, the total spin S is zero too, in the great majority of cases, in the normal state. In other words, the ground term of the molecule is ${}^1\Sigma^+$, and it is ${}^1\Sigma^+_g$ if the molecule consists of two similar atoms. Exceptions to these rules are formed by the molecules O_2 (whose normal term is ${}^3\Sigma_g^-$) and NO (normal term ${}^2\Pi$).

§57. The intersection of electron terms

The electron terms in a diatomic molecule as functions of the distance r between the nuclei can be represented graphically by plotting the energy as a function of r. It is of considerable interest to examine the intersection of the curves representing the different terms.

Let $U_1(r)$, $U_2(r)$ be two different electron terms. If they intersect at some point, then the functions U_1 and U_2 will have neighbouring values near this point. To decide whether such an intersection can occur, it is convenient to put the problem as follows. Let us consider a point r_0 where the functions $U_1(r)$, $U_2(r)$ have very close but not equal values (which we denote by E_1, E_2), and examine whether or not we can make U_1 and U_2 equal by displacing the point a short distance δr. The energies E_1 and E_2 are eigenvalues of the Hamiltonian $\hat{H}_0$ of the system of electrons in the field of the nuclei, which are at a distance r_0 from each other. If we add to the distance r_0 an increment δr, the Hamiltonian becomes $\hat{H}_0+\hat{V}$, where $\hat{V} = \delta r \cdot \partial \hat{H}_0/\partial r$ is a small correction; the values of the functions U_1, U_2 at the point $r_0+\delta r$ can be regarded as eigenvalues of the new Hamiltonian. This point of view enables us to determine the values of the terms $U_1(r)$, $U_2(r)$ at the point $r_0+\delta r$ by means of perturbation theory, $\hat{V}$ being regarded as a perturbation to the operator $\hat{H}_0$.

The ordinary method of perturbation theory is here inapplicable, however, since the eigenvalues E_1, E_2 of the energy in the unperturbed problem are very close to each other, and their difference is in general small compared with the magnitude of the perturbation; the condition (32.9) is not fulfilled. Since, in the limit as the difference E_2-E_1 tends to zero, we have the case of degenerate eigenvalues, it is natural to attempt to apply to the case of close eigenvalues a method similar to that developed in §33.

Let ψ_1, ψ_2 be the eigenfunctions of the unperturbed operator $\hat{H}_0$ which correspond to the energies E_1, E_2. As an initial zero-order approximation we take, instead of ψ_1 and ψ_2 themselves, linear combinations of them of the form

$$\psi = c_1\psi_1+c_2\psi_2. \tag{57.1}$$

Substituting this expression in the perturbed equation

$$(\hat{H}_0+\hat{V})\psi = E\psi, \tag{57.2}$$

we obtain

$$c_1(E_1+\hat{V}-E)\psi_1+c_2(E_2+\hat{V}-E)\psi_2 = 0.$$

Multiplying this equation on the left by ψ^* and ψ^* in turn, and integrating, we have two algebraic equations:

$$c_1(E_1+V_{11}-E)+c_2V_{12} = 0,$$
$$c_1V_{21}+c_2(E_2+V_{22}-E) = 0,$$

where $V_{ik} = \int \psi_i^* \hat{V} \psi_k \, \mathrm{d}q$. Since the operator $\hat{V}$ is Hermitian, the quantities V_{11} and V_{22} are real, while $V_{12} = V_{21}^*$. The compatibility condition for these equations is

$$\begin{vmatrix} E_1+V_{11}-E & V_{12} \\ V_{21} & E_2+V_{22}-E \end{vmatrix} = 0,$$

whence

$$E = \tfrac{1}{2}(E_1+E_2+V_{11}+V_{22}) \pm \sqrt{[\tfrac{1}{4}(E_1-E_2+V_{11}-V_{22})^2+|V_{12}|^2]}. \tag{57.3}$$

This formula gives the required eigenvalues of the energy in the first approximation.

If the energy values of the two terms become equal at the point $r_0+\delta r$ (i.e. the terms intersect), this means that the two values of E given by formula (57.3) are the same. For this to happen, the expression under the radical in (57.3) must vanish. Since it is the sum of two squares, we obtain, as the condition for there to be points of intersection of the terms, the equations

$$E_1-E_2+V_{11}-V_{22} = 0, \qquad V_{12} = 0. \tag{57.4}$$

However, we have at our disposal only one arbitrary parameter giving the perturbation $\hat{V}$, namely the magnitude δr of the displacement. Hence the two equations (57.4) cannot in general be simultaneously satisfied (we assume ψ_1 and ψ_2 to be chosen real and therefore V_{12} real).

It may happen, however, that the matrix element V_{12} vanishes identically; there then remains only one equation (57.4), which can be satisfied by a suitable choice of δr. This happens in all cases where the two terms considered are of different symmetry. By *symmetry* we here understand all possible forms of symmetry: with respect to rotations about an axis, reflections in planes, inversion, and also with respect to interchanges of electrons. In the diatomic molecule this means that we may be dealing with terms of different Λ, different parity or multiplicity, or (for Σ terms) Σ^+ and Σ^- terms.

This result arises because the operator $\hat{V}$ (like the Hamiltonian itself) commutes with all the symmetry operators for the molecule: the operator of the angular momentum about an axis, the reflection and inversion operators, and the operators of interchanges of electrons. It has been shown in §§18 and 19 that, for a scalar quantity whose operator commutes with the angular momentum and inversion operators, only the matrix elements for transitions between states of the same angular momentum and parity are non-zero. This proof remains valid, in essentially the same form, for the general case of an arbitrary symmetry operator.

Thus we reach the result that, in a diatomic molecule, only terms of different symmetry can intersect, while the intersection of terms of like symmetry is impossible (E. Wigner and J. von Neumann 1929).

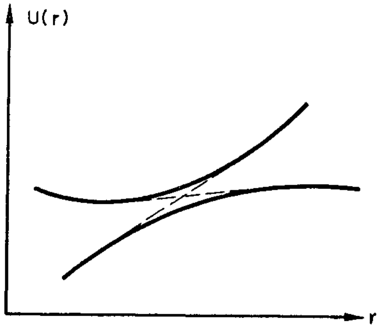

FIG. 12

If, as a result of some approximate calculation, we obtain two intersecting terms of the same symmetry, they are found to move apart on calculating the next approximation, as shown by the continuous lines in Fig. 12.

§58. Valency

The property of atoms of combining with one another to form molecules is described by means of the concept of *valency*. To each atom we ascribe a definite valency, and when atoms combine their valencies must be mutually satisfied, i.e. to each valency bond of an atom there must correspond a valency bond of another atom. For example, in the methane molecule CH_4, the four valency bonds of the quadrivalent carbon atom are satisfied by the four univalent hydrogen atoms. In going on to give a physical interpretation of valency, we shall begin with the simplest example, the combination of two hydrogen atoms to form the molecule H_2.

Let us consider two hydrogen atoms in the ground state (^2S). When they approach, the resulting system may be in the molecular state $^1\Sigma_g^+$ or $^3\Sigma_u^+$. The singlet term corresponds to an antisymmetrical spin wave function, and the triplet term to a symmetrical function. The coordinate wave function, on the other hand, is symmetrical for the $^1\Sigma$ term and antisymmetrical for the $^3\Sigma$ term. It is evident that the ground term of the H_2 molecule can only be the $^1\Sigma$ term. For an antisymmetrical wave function $\phi(\mathbf{r}_1, \mathbf{r}_2)$ (where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the radius vectors of the two electrons) always has nodes (since it vanishes for $\mathbf{r}_1 = \mathbf{r}_2$), and hence cannot belong to the lowest state of the system. A numerical calculation shows that the electron term $^1\Sigma$ in fact has a deep minimum corresponding to the formation of a stable H_2 molecule. In the $^3\Sigma$ state, the energy $U(\mathbf{r})$ decreases monotonically as the distance between the nuclei increases, corresponding to the mutual repulsion of the two hydrogen atoms[†] (Fig. 13).

Thus, in the ground state, the total spin of the hydrogen molecule is zero, $S = 0$. It is found that the molecules of practically all chemically stable compounds of elements of the principal groups have this property. Among inorganic molecules, exceptions are formed by the diatomic molecules O_2 (ground state $^3\Sigma$) and NO (ground state $^2\Pi$) and the triatomic molecules NO_2, ClO_2 (total spin $S = \frac{1}{2}$). Elements of the intermediate groups have special properties which we shall

---

[†] Here we ignore the van der Waals attraction forces between the atoms (see §61). The existence of these forces causes a minimum (at a greater distance) on the $U(r)$ curve for the $^3\Sigma$ term also. This minimum, however, is very shallow in comparison with that on the $^1\Sigma$ curve, and would not be perceptible on the scale of Fig. 13.

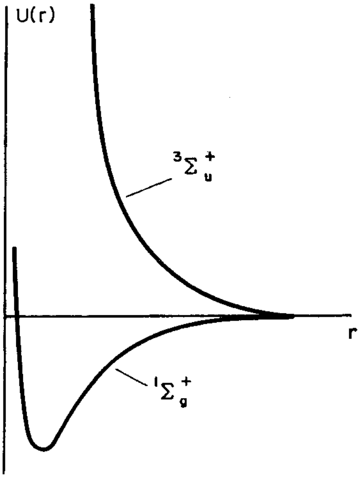

FIG. 13

discuss below, after studying the valency properties of the elements of the principal groups.

The property of atoms of combining with one another is thus related to their spin (W. Heitler and H. London 1927). The combination occurs in such a way that the spins of the atoms compensate one another. As a quantitative characteristic of the mutual combining powers of atoms, it is convenient to use an integer, twice the spin of the atom. This is equal to the chemical valency of the atom. Here it must be borne in mind that the same atom may have different valencies according to the state it is in.

Let us examine, from this point of view, the elements of the principal groups in the periodic system. The elements of the first group (the first column in Table 1 (§52), the group of alkali metals) have a spin $S = \frac{1}{2}$ in the normal state, and accordingly their valencies are unity. An excited state with a higher spin can be attained only by exciting an electron from a completed shell. Accordingly, these states are so high that the excited atom cannot form a stable molecule.

The atoms of elements in the second group (the second column in Table 1, the group of alkaline-earth metals) have a spin $S = 0$ in the normal state. Hence these atoms cannot enter into chemical compounds in the normal state. However, comparatively close to the ground state there is an excited state having a configuration sp instead of s^2 in the incomplete shell, and a total spin $S = 1$. The valency of an atom in this state is 2, and this is the principal valency of the elements in the second group.

The elements of the third group have an electron configuration s^2p in the normal state, with a spin $S = \frac{1}{2}$. However, by exciting an electron from the completed s-shell, an excited state is obtained having a configuration sp^2 and a spin $S = 3/2$, and this state lies close to the normal one. Accordingly, the elements of this group are both univalent and tervalent. The first two elements in the group (boron, aluminium) behave only as tervalent elements. The tendency to exhibit a valency 1 increases with the atomic number, and thallium behaves equally as a univalent and as a tervalent element (for example, in the compounds $TlCl$ and $TlCl_3$). This is due to the fact that, in the first few elements, the binding energy in the tervalent compounds is greater than for the univalent compounds, and this difference exceeds the excitation energy of the atom.

In the elements of the fourth group, the ground state has the configuration s^2p^2 with a spin of 1, and the adjacent excited state has a configuration sp^3 with a spin 2. The valencies 2 and 4 correspond to these states. As in the third group, the first two elements (carbon, silicon) exhibit mainly the higher valency (though the compound CO, for example, forms an exception), and the tendency to exhibit the lower valency increases with the atomic number.

In the atoms of the elements of the fifth group, the ground state has the configuration s^2p^3 with a spin $S = 3/2$, so that the corresponding valency is three. An excited state of higher spin can be obtained only by the transition of one of the electrons into the shell with the next higher value of the principal quantum number. The nearest such state has the configuration sp^3s' and a spin $S = 5/2$ (by s' we conventionally denote here an s state of an electron with a principal quantum number one greater than in the state s). Although the excitation energy of this state is comparatively high, the excited atom can still form a stable compound. Accordingly, the elements of the fifth group behave as both tervalent and quinquevalent elements (thus, nitrogen is tervalent in NH_3 and quinquevalent in HNO_3).

In the sixth group of elements, the spin is 1 in the ground state (configuration s^2p^4), so that the atom is bivalent. The excitation of one of the p electrons leads to a state s^2p^3s' of spin 2, while the excitation of an s electron in addition gives a state $sp^3s'p'$ of spin 3. In both excited states the atom can enter into stable molecules, and accordingly exhibits valencies of 4 and 6. The first element of the sixth group (oxygen) shows only valency 2, while the subsequent elements show higher valencies also (thus, sulphur in H_2S, SO_2, SO_3 is respectively bivalent, quadrivalent and sexivalent).

In the seventh group (the halogen group), the atoms are univalent in the ground state (configuration s^2p^5, spin $S = \frac{1}{2}$). They can, however, also enter into stable compounds when they are in excited states having configurations s^2p^4s', $s^2ps'^3p'$, $sp^3s'p'^2$ with spins 3/2, 5/2, 7/2 and valencies 3, 5, 7 respectively. The first element in the group (fluorine) is always univalent, but the subsequent elements also exhibit the higher valencies (thus, chlorine in HCl, $HClO_2$, $HClO_3$, $HClO_4$ is respectively univalent, tervalent, quinquevalent and septivalent).

Finally, the atoms of the elements in the group of inert gases have completely filled shells in their ground states (so that the spin $S = 0$), and their excitation energies are high. Accordingly, the valency is zero, and these elements are chemically inactive.

When atoms combine to form a molecule, the completed electron shells in the atoms are not much changed. The distribution of the electron density in the incomplete shells, on the other hand, may be considerably altered. In the most clearly defined cases of what is called *heteropolar binding*, all the valency electrons pass over from their own atoms to other atoms, so that we may say that the molecule consists of ions with charges equal (in units of e) to the valency. The elements of the first group are electropositive: in heteropolar compounds they lose electrons, forming positive ions. As we pass to the subsequent groups the electropositive character of the elements becomes gradually less marked and changes into electronegative character, which is present to the greatest extent in the elements of the seventh group. Regarding heteropolarity, however, the following remark should be made. If a molecule is heteropolar, this does not mean that, on moving the atoms apart, we necessarily obtain two ions. Thus, from the molecule CsF we should in fact obtain the ions Cs^+ and F^-, but the molecule NaF gives in the limit the neutral atoms Na and F (since the affinity of fluorine for an electron is greater than the ionisation potential of caesium but less than that of sodium).

In the opposite limiting case of what is called *homopolar binding*, the atoms in the molecule remain neutral on the average. Homopolar molecules, unlike heteropolar ones, have no appreciable dipole moment. The difference between the heteropolar and homopolar types is purely quantitative, and any intermediate case may occur.

Let us now turn to the elements of the intermediate groups. Those of the palladium and platinum groups are very similar to the elements of the principal groups as regards their valency properties. The only difference is that, owing to the comparatively deep position of the d electrons inside the atom, they interact only slightly with the other atoms in the molecule. As a result, "unsaturated" compounds, whose molecules have non-zero spin (though in practice not exceeding $\frac{1}{2}$), are often found among the compounds of these elements. Each of the elements can exhibit various valencies, and these may differ by unity, and not only by two as with the elements of the principal groups (where the change in valency is due to the excitation of some electron whose spin is compensated, so that the spins of two electrons are simultaneously released).

The elements of the rare-earth group are characterised by the presence of an incomplete f shell. The f electrons lie much deeper than the d electrons, and therefore take no part in the valency. Thus the valency of the rare-earth elements is determined only by the s and p electrons in the incomplete shells.[†] However, it must be borne in mind that, when the atom is excited, f electrons may pass into s and p states, thereby increasing the valency by one. Hence the rare-earth elements

---

[†] The d electrons which are found in the incomplete shells of the atoms of some rare-earth elements are unimportant, since these atoms in practice always form compounds in excited states where there are no d electrons.

too exhibit valencies differing by unity (in practice they are all tervalent and quadrivalent).

The elements of the iron group occupy, as regards their valency properties, a position intermediate between the rare-earth elements and those of the palladium and platinum groups. In their atoms, the *d* electrons lie comparatively deep, and in many compounds take no part in the valency bonds. In these compounds, therefore, the elements of the iron group behave like rare-earth elements. Such compounds include those of ionic type (for instance $FeCl_2$, $FeCl_3$), in which the metal atom enters as a simple cation. Like the rare-earth elements, the elements of the iron group can show very various valencies in these compounds.

Another type of compound of the iron-group elements is formed by what are called *complex compounds*. These are characterised by the fact that the atom of the intermediate element enters into the molecule not as a simple ion, but as part of a complex ion (for instance the ion MnO_4^- in $KMnO_4$, or the ion $Fe(CN)_6^{4-}$ in $K_4Fe(CN)_6$). In these complex ions, the atoms are closer together than in simple ionic compounds, and in them the *d* electrons take part in the valency bond. Accordingly, the elements of the iron group behave in complex compounds like those of the palladium and platinum groups.

Finally, it must be mentioned that the elements copper, silver and gold, which in §52 we placed among the principal groups, behave as intermediate elements in some of their compounds. These elements can exhibit valencies of more than one, on account of a transition of an electron from a *d* shell to a *p* shell of nearly the same energy (for example, from 3*d* to 4*p* in copper). In such compounds the atoms have an incomplete *d* shell, and hence behave as intermediate elements: copper like the elements of the iron group, and silver and gold like those of the palladium and platinum groups.

§59. Vibrational and rotational structures of terms in the diatomic molecule

As has been pointed out at the beginning of this chapter, the great difference in the masses of the nuclei and the electrons makes it possible to divide the problem of determining the energy levels of a molecule into two parts. We first determine the energy levels of the system of electrons, for nuclei at rest, as functions of the distance between the nuclei (the electron terms). We can then consider the motion of the nuclei for a given electron state; this amounts to regarding the nuclei as particles interacting with one another in accordance with the law $U_n(r)$, where U_n is the corresponding electron term. The motion of the molecule is composed of its translational displacement as a whole, together with the motion of the nuclei about their centre of mass. The translational motion is, of course, without interest, and we can regard the centre of mass as fixed.

We shall consider only the electron terms in which the total spin S of the molecule is zero (the singlet terms). This simple case presents all the principal qualitative features of the structure of energy levels in the diatomic molecule.

The problem of the relative motion of two particles (the nuclei) which interact according to a law $U(r)$ depending only on the distance r between them reduces to that of the motion of a single particle of mass M (the reduced mass of the two particles) in a central field $U(r)$. This reduces in turn to that of a one-dimensional motion in a field where the effective potential energy is equal to the sum of U and the centrifugal energy (cf. §29).

When the spin is zero, the total angular momentum $\mathbf{J}$ of the molecule consists of the orbital angular momentum $\mathbf{L}$ of the electrons and the angular momentum of rotation of the nuclei. The latter angular momentum therefore corresponds to the operator $\hat{\mathbf{J}}-\hat{\mathbf{L}}$, and the centrifugal-energy operator is

$$\frac{\hbar^2}{2Mr^2}(\hat{\mathbf{J}}-\hat{\mathbf{L}})^2.$$

The effective potential energy is defined as

$$U_J(r) = U(r)+\frac{\hbar^2}{2Mr^2}\overline{(\mathbf{J}-\mathbf{L})^2}, \tag{59.1}$$

where the averaging is taken with respect to the state of the molecule for a fixed value of r.

Let us carry out this averaging for a state in which the molecule has a definite value of the squared total angular momentum $\mathbf{J}^2 = J(J+1)$ and a definite value $L_z = \Lambda$ of the projection of the electron angular momentum on the axis of the molecule (the z-axis). Expanding the square in (59.1) gives

$$U_J(r) = U(r)+\frac{\hbar^2}{2Mr^2}J(J+1)-\frac{\hbar^2}{Mr^2}\overline{\mathbf{L}}\,.\,\mathbf{J}+\frac{\hbar^2}{2Mr^2}\overline{\mathbf{L}^2}. \tag{59.2}$$

The last term depends only on the electron state, and does not contain the quantum number J; it can be included in the energy $U(r)$. We shall show that the same is true of the penultimate term.

If the projection of the angular momentum on some axis has a definite value, the mean value of the angular momentum vector is directed along that axis (see the end of §15). Thus, if $\mathbf{n}$ denotes a unit vector along the z-axis, $\overline{\mathbf{L}} = \Lambda\mathbf{n}$. In classical mechanics, the angular momentum of rotation of a system of two particles (the nuclei) is $\mathbf{r}\times\mathbf{p}$, where $\mathbf{r} = r\mathbf{n}$ is the radius vector joining the two particles and $\mathbf{p}$ is the momentum of their relative motion; this vector is perpendicular to $\mathbf{n}$. In quantum mechanics, the same is true of the rotational angular momentum operator: $(\hat{\mathbf{J}}-\hat{\mathbf{L}})\,.\,\mathbf{n} = 0$, or $\hat{\mathbf{J}}\,.\,\mathbf{n} = \hat{\mathbf{L}}\,.\,\mathbf{n}$. Since the operators are equal, so are their eigenvalues; since $\mathbf{n}\,.\,\mathbf{L} = L_z = \Lambda$, it follows that

$$J_z = \Lambda. \tag{59.3}$$

Thus $\overline{\mathbf{L}}\,.\,\mathbf{J}$ in the penultimate term in (59.2) is equal to $\mathbf{n}\,.\,\mathbf{J}\Lambda = \Lambda^2$, i.e. is independent of J. Redefining the function $U(r)$, we can write the effective potential energy in the form

$$U_J(r) = U(r)+\frac{\hbar^2}{2Mr^2}J(J+1). \tag{59.4}$$

On solving the one-dimensional Schrödinger's equation with this potential energy, we obtain a series of energy levels. We arbitrarily number these levels (for each given J) in order of increasing energy, using a number $v = 0, 1, 2, \ldots$; $v = 0$ corresponds to the lowest level. Thus the motion of the nuclei causes a splitting of each electron term into a series of levels characterised by the values of the two quantum numbers J and v.

The dependence of the energy levels on the quantum numbers cannot be completely calculated in a general form. Such a calculation is possible only for low excited levels which lie not too far above the ground level. Small values of the quantum numbers J and v correspond to these levels. It is with such levels that we are in fact most often concerned in the study of molecular spectra, and hence they are of particular interest.

The motion of the nuclei in slightly excited states can be regarded as small vibrations about the equilibrium position. Accordingly we can expand $U(r)$ in a series of powers of $\xi = r - r_e$, where r_e is the value of r for which $U(r)$ has a minimum. Since $U'(r_e) = 0$, we have as far as terms of the second order

$$U(r) = U_e + \tfrac{1}{2}M\omega^2\xi^2,$$

where $U_e = U(r_e)$, and $\omega = \sqrt{[U''(r_e)/M]}$ is the frequency of the vibrations (see *Mechanics and Electrodynamics*, §17).

In the second term in (59.4)—the centrifugal energy— it is sufficient to put $r = r_e$. Thus we have

$$U_J(r) = U_e + BJ(J+1) + \tfrac{1}{2}M\omega^2\xi^2, \tag{59.5}$$

where $B = \hbar^2/2Mr_e^2 = \hbar^2/2I$ is what is called the *rotational constant* ($I = Mr_e^2$ is the moment of inertia of the molecule).

The first two terms in (59.5) are constants, while the third corresponds to a one-dimensional harmonic oscillator. Hence we can at once write down the required energy levels:

$$E = U_e + BJ(J+1) + \hbar\omega(v + \tfrac{1}{2}). \tag{59.6}$$

Thus, in the approximation considered, the energy levels are composed of three independent parts:

$$E = E_{el} + E_r + E_v. \tag{59.7}$$

The term $E_{el} = U_e$ is the electron energy (including the energy of the Coulomb interaction of the nuclei); the second term

$$E_r = BJ(J+1) \tag{59.8}$$

is the rotational energy from the rotation of the molecule.[†] Since the projection of the angular momentum cannot exceed its magnitude J, it follows from (59.3) that the quantum number J can take only the values

$$J = \Lambda, \quad \Lambda+1, \quad \Lambda+2, \ldots \,. \tag{59.9}$$

Finally, the third term in (59.7)

$$E_v = \hbar\omega(v+\tfrac{1}{2}) \tag{59.10}$$

is the energy of the vibrations of the nuclei within the molecule. The number v denumerates, by definition, the levels with a given J in order of increasing energy; it is called the *vibrational quantum number.*

For a given form of the potential energy curve $U(r)$, the frequency ω is inversely proportional to $\sqrt{M}$. Hence the intervals ΔE_v between the vibrational levels are proportional to $1/\sqrt{M}$. The intervals ΔE_r between the rotational levels contain in the denominator the moment of inertia I, and are therefore proportional to $1/M$. The intervals ΔE_{el} between the electron levels, however, are independent of M, like the levels themselves. Since m/M (m being the electron mass) is a small parameter in the theory of diatomic molecules, we see that

$$\Delta E_{el} \gg \Delta E_v \gg \Delta E_r. \tag{59.11}$$

These inequalities show that the distribution of the energy levels of the molecule is rather unusual. The vibrational motion of the nuclei splits the electron terms into levels lying comparatively close together.

[†] A rotating system of two rigidly connected particles is often called a *rotator.* Formula (59.8) gives the quantum energy levels of a rotator.

These levels, in turn, exhibit a fine splitting due to the rotational motion of the molecule. As an example, we give the values of U_e, $\hbar\omega$, and B (in electron-volts) for a few typical molecules:

	H_2	N_2	O_2
$-U_e$	4.7	7.5	5.2
$\hbar\omega$	0.54	0.29	0.20
$10^3\times B$	7.6	0.25	0.18

§60. Parahydrogen and orthohydrogen

In §56 we have already examined some symmetry properties of the states of a diatomic molecule. These pertained to the electron terms, i.e. characterised the behaviour of the electron wave function in transformations not affecting the coordinates of the nuclei. When the motion of the nuclei (vibration and rotation), is included in the state of the molecule, new symmetry properties appear which relate to the molecule as a whole. Here we shall discuss an interesting effect arising from the symmetry of states of diatomic molecules that consist of like atoms (belonging not only to the same element but to the same isotope, so that the two nuclei are identical), and take the particular case of a hydrogen molecule in its electron ground state (the singlet state $^1\Sigma_g^+$).

The Hamiltonian of a molecule of like atoms is invariant with respect to an interchange of the nuclei. There is consequently a new symmetry property of the states: the wave function of the molecule may be symmetric or antisymmetric with respect to a change in the sign of the radius vector **r** from one nucleus to the other.

The wave function of the molecule is the product of the electron and nuclear wave functions. It has been shown in §59 that the latter is formally identical with the wave function of a single particle with orbital angular momentum J in a centrally symmetric field $U(r)$. From this point of view, the transformation $\mathbf{r} \rightarrow -\mathbf{r}$ is an inversion of the coordinates with respect to the centre of the field, and according to (19.5) such a transformation multiplies the wave function by $(-1)^J$.

The electron wave function also depends on the coordinates of the nuclei as parameters. For the electron ground term of the molecule, this function is symmetrical with respect to interchange of the nuclei.[†] Hence the factor $(-1)^J$ determines the symmetry or antisymmetry not only of the nuclear part but also of the entire wave function of the molecule.

In §46 we have established a general theorem that, for a system of two like particles with spin $i = \frac{1}{2}$, the states symmetrical with respect to the coordinates of the particles can occur only for a zero total spin I of the particles, and the antisymmetric states only for $I = 1$. This rule can be applied to the two nuclei in the hydrogen molecule (protons with spin $\frac{1}{2}$), and the result is that, when the nuclear spins are parallel ($I = 1$) the molecule in its normal electron state must have an odd rotational angular momentum J; when the nuclear spins are antiparallel ($I = 0$), J must be even. This is a noteworthy instance of the quantum-mechanical exchange effect: the nuclear spins exert a strong indirect influence on the molecular terms, although their direct influence on the energy (the hyperfine structure of the terms) is entirely negligible.

Since the magnetic moments of the protons are extremely small, and hence their spins interact only weakly with the electrons in the molecule, the probability of a change in I is very small even in collisions between molecules. Hence the molecules with $I = 1$ and $I = 0$ behave almost as different forms of matter; they are referred to as *orthohydrogen* and *parahydrogen* molecules respectively.

The ground level of the parahydrogen molecule corresponds to the rotational quantum number $J = 0$. For the orthohydrogen molecule, which can have only odd values of J, the ground level is $J = 1$, which is higher than the ground level of parahydrogen.

[†] This property corresponds to the general empirical rule stated in §56, whereby in the majority of diatomic molecules the normal electron state is completely symmetrical. It can also be shown directly that the symmetry with respect to interchange of the nuclei follows automatically from the other properties of the ${}^1\Sigma_g^+$ state: the symmetry under reflection in a plane through the axis of the molecule and under change of sign of all the electron coordinates, the coordinates of the nuclei remaining unchanged.

§61. Van der Waals forces

Let us consider two atoms which are at a great distance from each other (relative to their size), and determine the energy of their interaction. In other words, we shall discuss the determination of the form of the electron terms $U_n(r)$ when the distance between the nuclei is large.

To solve this problem we apply perturbation theory, regarding the two isolated atoms as the unperturbed system, and the potential energy of their electrical interaction as the perturbation operator. As we know (see *Mechanics and Electrodynamics*, §64), the electrical interaction of two systems of charges at a large distance r apart can be expanded in powers of $1/r$, and successive terms of this expansion correspond to the interaction of the total charges, dipole moments, quadrupole moments, etc., of the two systems. For neutral atoms, the total charges are zero. The expansion here begins with the dipole–dipole interaction ($\sim 1/r^3$); then follow the dipole–quadrupole terms ($\sim 1/r^4$), the quadrupole–quadrupole terms ($\sim 1/r^5$), and so on.

Let us first suppose that both atoms are in the S state. Then it is easily seen that there is no interaction between the atoms in the first approximation of perturbation theory. The energy of the interaction of the atoms is there determined as the diagonal matrix element of the perturbation operator, calculated with respect to the unperturbed wave functions of the system (expressed in terms of products of the wave functions for the two atoms). In S states, however, the diagonal matrix elements, i.e. the mean values of the dipole, quadrupole, etc. moments, are zero; this follows since the distribution of charges in the atoms is spherically symmetrical on the average.

In the second approximation it is sufficient to restrict ourselves to the dipole interaction in the perturbation operator, since this decreases least rapidly as r increases, i.e. to the term

$$V = [-\mathbf{d}_1 \cdot \mathbf{d}_2 + 3(\mathbf{d}_1 \cdot \mathbf{n})(\mathbf{d}_2 \cdot \mathbf{n})]/r^3, \qquad (61.1)$$

where $\mathbf{n}$ is a unit vector in the direction joining the two atoms. Since the non-diagonal matrix elements of the dipole moment are in general different from zero, we obtain in the second approximation of perturbation theory a non-vanishing result which, being quadratic in V, is proportional to $1/r^6$. The correction in the second approximation to the lowest eigenvalue is always negative (§32). Hence we obtain for the interaction energy of atoms in their normal states an expression of the form

$$U(r) = -\text{constant}/r^6, \tag{61.2}$$

where the constant is positive (F. London 1928).

Thus two atoms in normal S states, at a great distance apart, attract each other with a force $(-\mathrm{d}U/\mathrm{d}r)$ which is inversely proportional to the seventh power of the distance. The attractive forces between atoms at large distances are usually called *van der Waals forces*. These forces cause the appearance of minima on the potential energy curves of the electron terms even for atoms which do not form a stable molecule. These depressions, however, are very shallow (being only tenths or even hundredths of an electron-volt in depth) and lie at distances several times greater than the distances between atoms in stable molecules.

Formula (61.2) is also important because it represents the interaction forces at large distances between atoms in any normal (not necessarily S) states, provided that this interaction is averaged over all possible orientations of the atoms; the interaction of atoms in a gas, for example, is of this type.[†] Although the mean dipole moment is zero in any stationary state, the mean value of the quadrupole moment may be non-zero for an atom with a non-zero angular momentum J (§54). The quadrupole–quadrupole term in the interaction operator may therefore yield a non-zero result in the first approximation of perturbation theory. But the mean values of the quadrupole moment (and of the higher-order multipole moments) depend on the orientation of its angular momentum $\mathbf{J}$, and vanish on averaging with respect to this orientation, by symmetry.

[†] The law derived above on the basis of non-relativistic theory is, however, valid only so long as the retardation of electromagnetic interactions is unimportant. For this to be so, the distance r between the atoms must be small compared with c/ω_{0n}, where ω_{0n} are the frequencies of transitions between the ground state and the excited states of the atom.

PROBLEM

Derive a formula giving the van der Waals forces in terms of the matrix elements of dipole moments for two like atoms in S states.

Solution. The answer is obtained by applying the general formula (32.10) of perturbation theory to the operator (61.1). On account of the isotropy of the atoms in the S state it is evident *a priori* that, on summation over all intermediate states, the squared matrix elements of the three components of each of the vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ give equal contributions, while the terms which contain products of different components give zero. The result is

$$U(r) = -\frac{6}{r^6}\sum_{n,n'}\frac{(d_z)^2_{0n}(d_z)^2_{0n'}}{2E_0-E_n-E_{n'}},$$

where E_0 and E_n are the unperturbed values of the energies of the ground state and excited states of the atom.

CHAPTER 9

ELASTIC COLLISIONS

§62. The scattering amplitude

In classical mechanics, collisions of two particles are entirely determined by their velocities and impact parameter (i.e. the distance at which they would pass if they did not interact). In quantum mechanics the very wording of the problem must be changed, since in motion with definite velocities the concept of the path is meaningless, and therefore so is the impact parameter. The purpose of the theory is here only to calculate the probability that, as a result of the collision, the particles will deviate (or, as we say, be *scattered*) through any given angle. We are speaking here of what are called *elastic collisions*, in which the particles, or the internal state of the colliding particles if these are complex, are left unchanged.

The problem of an elastic collision, like any problem of two bodies, amounts to a problem of the scattering of a single particle, with the reduced mass, in the field $U(r)$ of a fixed centre of force.† This simplification is effected by changing to a system of coordinates in which the centre of mass of the two particles is at rest. The scattering angle in this system we denote by θ. It is simply related to the angles ϑ_1 and ϑ_2 giving the deviations of the two particles in the laboratory system of coordinates, in which the second particle (say) was at rest before

---

† Here we neglect the spin–orbit interaction of the particles (if they have spin). By assuming the field to be centrally symmetric, we exclude from consideration also processes such as the scattering of electrons by molecules.

the collision:

$$\tan \vartheta_1 = m_2 \sin \theta / (m_1 + m_2 \cos \theta), \quad \vartheta_2 = \tfrac{1}{2}(\pi - \theta), \tag{62.1}$$

where m_1, m_2 are the masses of the particles (see *Mechanics and Electrodynamics*, §14). In particular, if the masses of the two particles are the same ($m_1 = m_2$), we have simply

$$\vartheta_1 = \tfrac{1}{2}\theta, \quad \vartheta_2 = \tfrac{1}{2}(\pi - \theta); \tag{62.2}$$

the sum $\vartheta_1 + \vartheta_2 = \frac{1}{2}\pi$, i.e. the particles diverge at right angles.

In what follows, we shall always use (unless the contrary is specifically stated) a system of coordinates in which the centre of mass is at rest, and m will denote the reduced mass of the colliding particles.

A free particle moving in the positive direction of the z-axis is described by a plane wave, which we take in the form $\psi = e^{ikz}$, i.e. the current density in the wave is equal to the particle velocity v; cf. the normalisation to unit current in (21.6). The scattered particles must be described, at a great distance from the scattering centre, by an outgoing spherical wave of the form $f(\theta)e^{ikr}/r$, where $f(\theta)$ is some function of the scattering angle θ (the angle between the z-axis and the direction of the scattered particle).[†] Thus the solution of Schrödinger's equation for a scattering process in a field $U(r)$ must have at large distances the asymptotic form

$$\psi \approx e^{ikz} + f(\theta)e^{ikr}/r. \tag{62.3}$$

The function $f(\theta)$ is called the *scattering amplitude*. The probability per unit time that the scattered particle will pass through a surface element $\mathrm{d}S = r^2\,\mathrm{d}o$ (where $\mathrm{d}o$ is an element of solid angle) is $(v/r^2)\,|f|^2\,\mathrm{d}S = v\,|f|^2\,\mathrm{d}o$.[‡] Its ratio to the current density in the

---

[†] An outgoing spherical wave contains an exponential factor e^{ikr}, and an ingoing wave contains a corresponding factor e^{-ikr}, in place of the trigonometric factor in the stationary spherical waves discussed in §30.

[‡] It is supposed that the incident beam of particles is defined by a wide (to avoid diffraction effects) but finite diaphragm, as happens in actual experiments on scattering. There is therefore no interference between the two terms of the expression (62.3); the squared modulus $|\psi|^2$ is taken at points where there is no incident wave.

incident wave is

$$d\sigma = |f(\theta)|^2 \, do. \tag{62.4}$$

This quantity has the dimensions of area, and is called the *effective cross-section*, or simply the *cross-section*, for scattering into the solid angle do. If we put $do = 2\pi \sin\theta \, d\theta$, we obtain for the cross-section

$$d\sigma = 2\pi \sin\theta \, |f(\theta)|^2 \, d\theta \tag{62.5}$$

for scattering through angles in the range from θ to $\theta + d\theta$.

A solution of Schrödinger's equation for scattering in a central field is axially symmetric about the z-axis. The general form of such a solution can be represented as the expansion

$$\psi = \sum_{l=0}^{\infty} A_l P_l(\cos\theta) R_{kl}(r), \tag{62.6}$$

where the R_{kl} are radial functions satisfying equation (29.8) (with energy $E = \hbar^2 k^2/2m$). The asymptotic form of these functions at large distances is given by the stationary waves (30.10). We shall show how the scattering amplitude may be expressed in terms of the phase shifts δ_l of these functions.

Substituting (30.10) in (62.6), we can write the general asymptotic form of the wave function as

$$\begin{aligned}\psi &\approx \sqrt{\frac{2}{\pi}} \frac{1}{r} \sum_{l=0}^{\infty} A_l P_l \sin(kr - \tfrac{1}{2} l\pi + \delta_l) \\ &= \frac{1}{2} i \sqrt{\frac{2}{\pi}} \frac{1}{r} \sum_{l=0}^{\infty} A_l P_l \{\exp[-i(kr - \tfrac{1}{2} l\pi + \delta_l)] \\ &\quad - \exp[i(kr - \tfrac{1}{2} l\pi + \delta_l)]\}.\end{aligned}$$

We have to choose the coefficients A_l such that this function has the form (62.3). To do this, we use the expansion of a plane wave in terms of spherical waves, obtained in §30. For large r, according to (30.16),

$$e^{ikz} \approx \frac{i}{2kr} \sum_{l=0}^{\infty} i^l (2l+1) P_l \{\exp[-i(kr - \tfrac{1}{2} l\pi)] - \exp[i(kr - \tfrac{1}{2} l\pi)]\}.$$

The difference $\psi - e^{ikz}$ must represent an outgoing wave, i.e. must contain no term in e^{-ikr}; thus

$$A_l = \frac{1}{k}\sqrt{\frac{\pi}{2}}(2l+1)i^l e^{i\delta_l}.$$

Hence

$$\psi \approx \frac{i}{2kr}\sum_{l=0}^{\infty}(2l+1)P_l(\cos\theta)[(-1)^l e^{-ikr} - e^{2i\delta_l}e^{ikr}], \qquad (62.7)$$

For the coefficient of e^{ikr}/r in the difference $\psi - e^{ikz}$, we obtain

$$f(\theta) = \frac{1}{2ik}\sum_{l=0}^{\infty}(2l+1)[e^{2i\delta_l}-1]P_l(\cos\theta). \qquad (62.8)$$

The formula solves the problem of expressing the scattering amplitude in terms of the δ_l (H. Faxén and J. Holtsmark 1927). Each term of the sum is called a *partial amplitude.*

If we integrate dσ over all angles, we obtain the total scattering cross-section σ, which is the ratio of the total probability (per unit time) that the particle will be scattered to the current density. Substituting (62.8) in the integral

$$\sigma = 2\pi\int_0^{\pi}|f(\theta)|^2\sin\theta\, d\theta,$$

and recalling that the polynomials $P_l(\cos\theta)$ are orthogonal, we see that only the squares of the individual terms in the sum (62.8) remain, and the normalisation integral (30.13) then gives

$$\sigma = \frac{4\pi}{k^2}\sum_{l=0}^{\infty}(2l+1)\sin^2\delta_l. \qquad (62.9)$$

§63. The condition for quasi-classical scattering

The limiting transition from the exact quantum-mechanical formulae of scattering theory, derived in §62, to the classical formulae is quite lengthy and will not be given here. Instead, some comments will be made about the conditions under which this transition is possible.

If we can speak of classical scattering through an angle θ when the particle is incident at an impact parameter ϱ, it is necessary that the quantum-mechanical indeterminacies of these two quantities should be relatively small: $\Delta\varrho \ll \varrho$, $\Delta\theta \ll \theta$. The indeterminacy in the scattering angle is of the order of magnitude $\Delta\theta \sim \Delta p/p$, where p is the momentum of the particle and Δp is the indeterminacy in its transverse component. Since $\Delta p \sim \hbar/\Delta\varrho \gg \hbar/\varrho$, we have $\Delta\theta \gg \hbar/p\varrho$, and thus

$$\theta \gg \hbar/\varrho m v. \tag{63.1}$$

Replacing the angular momentum $mv\varrho$ by $\hbar l$, we obtain $\theta l \gg 1$, and therefore $l \gg 1$, in accordance with the general rule that the quasi-classical case corresponds to large values of the quantum numbers (§27).

The classical angle of deviation of the particle can be estimated as the ratio of the transverse momentum increment Δp during the "collision time" $\tau \sim \varrho/v$ and the original momentum mv. The force acting on the particle at a distance ϱ in a field $U(r)$ is $F = -\mathrm{d}U(\varrho)/\mathrm{d}\varrho$; hence $\Delta p \sim F\varrho/v$, so that $\theta \sim \varrho F/mv^2$. This estimate is strictly valid only if $\theta \ll 1$, but it can be applied to give an order of magnitude even if $\theta \sim 1$. Substitution in (63.1) gives the condition for quasi-classical scattering in the form

$$F\varrho^2 \gg \hbar v. \tag{63.2}$$

If the field $U(r)$ decreases more rapidly than $1/r$, the condition (63.2) always ceases to be satisfied for sufficiently large ϱ. Small θ, however, correspond to large ϱ; thus scattering through sufficiently small angles is never classical. The quantum nature of the scattering through small angles is, in particular, the reason why the total scattering cross-section may be finite. In this connection it may be recalled that in classical mechanics, for any field which vanishes only as $r \to \infty$ (i.e. which is not sharply cut off at a finite distance), a particle passing at a large but finite impact parameter undergoes a deviation through a small but non-zero angle; the total cross-section, therefore, is always infinite. It is clear from the above discussion that in quantum mechanics the corresponding argument is invalid, since the concept of scattering ceases to be meaningful when the scattering angle is less than the quantum indeterminacy of the direction of motion of the particle.

§64. Discrete energy levels as poles of the scattering amplitude

There is a relation between the law of scattering of particles (with a positive energy E) in a given field and the discrete spectrum of negative energy levels (if any) in that field.

To simplify the formulae, we shall consider the motion of particles with orbital angular momentum $l = 0$. The asymptotic expression for the wave function with positive energy at a large distance from the centre of the field may be written as a sum of outgoing and ingoing spherical waves:

$$\psi = \frac{1}{r}\{a(k)e^{ikr}+b(k)e^{-ikr}\}. \tag{64.1}$$

The coefficients $a(k)$ and $b(k)$ are some functions of k, which could be determined only by solving Schrödinger's equation at short distances and using the finiteness of the wave function for $r = 0$. The two functions are then not independent, but are related in a simple manner. One relation follows immediately because the function ψ, being the wave function of a non-degenerate state, must be real:

$$b(k) = a^*(k). \tag{64.2}$$

If we next consider formally any values of k (real or complex), $a(k)$ and $b(k)$ become functions of a complex variable, still related by (64.2), and also by

$$a(-k) = b(k), \tag{64.3}$$

which follows from the definition of a and b in (64.1) (the coefficients a and b being interchanged when k is replaced by $-k$). The function ψ with complex k, being an analytical continuation of the solution of Schrödinger's equation with real k, is again the solution of this equation that is finite at the origin. It will not, however, satisfy the condition of finiteness in all space: when $r \to \infty$, one of the terms in (64.1) (depending on the sign of the imaginary part of k) becomes infinite.

In particular, when k is purely imaginary, the expression (64.1) gives the asymptotic form of the solution of Schrödinger's equation with a negative energy E. But, for this solution to correspond to a stationary state of the discrete spectrum, the function ψ must remain finite as $r \to \infty$. To each negative value of E there corresponds a pair of purely imaginary values $k = \pm i\sqrt{(2m|E|)}/\hbar$. With the upper sign, the second term in (64.1) does not satisfy the condition of finiteness as $r \to \infty$; thus, for a value of E corresponding to a discrete energy level, we must have

$$b(i|k|) = 0. \tag{64.4}$$

Similarly, when $k = -i|k|$, the function $a(k)$ must vanish.

Now, comparing (64.1) with the asymptotic expression for the wave function of a particle with energy $E > 0$ in the form (30.10)

$$\psi = \sqrt{\frac{2}{\pi}}\,\frac{1}{2ir}\left(e^{i(kr+\delta_0)} - e^{-i(kr+\delta_0)}\right),$$

we see that the ratio a/b is related to the phase δ_0 by

$$e^{2i\delta_0(k)} = a(k)/b(k). \tag{64.5}$$

This expression has a pole at the point where $b(k)$ is zero. The partial amplitude for s-wave scattering is

$$f_0 = \frac{1}{2ik}\left(e^{2i\delta_0} - 1\right),$$

and we therefore conclude that this amplitude as an analytic function of the complex variable k has poles in the upper half-plane of k with imaginary values of k corresponding to the energy levels of bound s states of the particle in the field.

A similar relation exists between the energy levels of bound states with $l \neq 0$ and the poles of the corresponding partial scattering amplitudes.

§65. The scattering of slow particles

Let us consider the properties of elastic scattering in the limiting case where the velocities of the particles undergoing scattering are so small that the de Broglie wavelength of a particle is large compared with the radius of action a of the scattering field[†], and the energy is small compared with the field within that radius: $ak \ll 1$ and $k^2\hbar^2/2m \ll |U|$.

The probability of finding the particle near the centre of the field (at a distance much less than the wavelength of the particle) decreases rapidly as the orbital angular momentum l increases (cf. the end of §29). Thus s-wave scattering ($l = 0$) plays the principal role in slow-particle scattering. In this case, to determine the properties of the scattering, we must find the limiting form of the dependence of the phase δ_0 on the wave number k when the latter is small.

The wave function for an s state depends only on r. When $r \lesssim a$ (within the radius of action of the field), in the exact Schrödinger's equation

$$\triangle\psi + k^2\psi = \frac{2m}{\hbar^2} U(r)\psi \tag{65.1}$$

we can neglect only the term in k^2:

$$\triangle\psi \equiv \frac{1}{r}(r\psi)'' = \frac{2m}{\hbar^2} U(r)\psi \qquad (r \lesssim a) \tag{65.2}$$

(the prime denoting differentiation with respect to r). At large distances, $a \ll r \ll 1/k$, we can also omit the term in $U(r)$, leaving

$$(r\psi)'' = 0. \tag{65.3}$$

The general solution of this equation is

$$\psi = c_1 + \frac{c_2}{r} \qquad (a \ll r \ll 1/k). \tag{65.4}$$

[†] The quantity a represents the linear dimension of the region of space in which the field U is significantly different from zero. For example, in the scattering of neutrons by nuclei, a is the nuclear radius; in the scattering of electrons by neutral atoms, it is the atomic radius.

The values of the real constants c_1 and c_2 can in principle be determined only by solving equation (65.2) for a particular function $U(r)$.

At still greater distances, with $r \gtrsim 1/k$, the term in $U(r)$ can be omitted from (65.1), but the term in k^2 cannot be neglected, so that we have

$$\frac{1}{r}(r\psi)''+k^2\psi = 0,$$

i.e. the equation of free motion. The solution of this equation is

$$\psi = \frac{c_1}{k}\,\frac{\sin kr}{r}+c_2\frac{\cos kr}{r} \qquad (r \gtrsim 1/k). \tag{65.5}$$

The coefficients have been chosen so that, for $kr \ll 1$, this solution becomes (65.4); this ensures the "joining" of the solutions in the regions $kr \ll 1$ and $kr \sim 1$.

We can write the sum (65.5) in the form

$$\psi = \frac{c_1}{kr}\sin(kr+\delta_0),$$

and obtain for the phase δ_0

$$\tan\delta_0 \approx \delta_0 = c_2k/c_1; \tag{65.6}$$

since k is small, the phase δ_0 is also small. Finally, retaining only the first term in (62.8), we find as the scattering amplitude

$$f = \frac{1}{2ik}(e^{2i\delta_0}-1) \approx \frac{\delta_0}{k} = \frac{c_2}{c_1}. \tag{65.7}$$

Thus the scattering amplitude is constant, depending neither on the scattering angle nor on the velocity of the particles. In other words, the scattering of slow particles is isotropic, and the cross-section $\sigma = 4\pi(c_2/c_1)^2$ is independent of the energy.†

---

† In the above discussion it has been tacitly assumed that the field $U(r)$ decreases sufficiently rapidly at large distances ($r \gg a$). It is easy to see just how rapid this decrease has to be. For large r, the second term in the function (65.4) is small in comparison with the first term. In order for the retention of the second term to be nevertheless legitimate, the small term $\sim c_2/r^3$ retained in (65.2) must still be large compared with the term $U\psi \sim Uc_1$ omitted in going from (65.2) to (65.3). Hence it follows that U must decrease more rapidly than $1/r^3$.

PROBLEMS

Problem 1. Determine the scattering amplitude for slow particles in a spherical potential well of depth U_0 and radius a: $U(r) = -U_0$ for $r < a$, $U(r) = 0$ for $r > a$.

Solution. The wave number of the particles is assumed to satisfy the conditions $ka \ll 1$ and $k \ll \varkappa$, where $\varkappa = \sqrt{(2mU_0)}/\hbar$. Equation (65.2) for the function $\chi = r\psi$ becomes

$$\chi'' + \varkappa^2\chi = 0$$

for $r < a$. The solution which makes ψ finite at $r = 0$ is

$$\chi = A \sin \varkappa r \qquad (r < a).$$

For $r > a$, the function χ satisfies the equation $\chi'' + k^2\chi = 0$, whence

$$\chi = B \sin (kr + \delta_0) \qquad (r > a).$$

From the continuity of χ'/χ at $r = a$, we obtain the equation

$$\varkappa \cot \varkappa a = k \cot (ka + \delta_0) \approx k/(ka + \delta_0),$$

from which we determine δ_0. As a result, we have for the scattering amplitude

$$f = \frac{\tan \varkappa a - \varkappa a}{\varkappa}. \tag{1}$$

If also $\varkappa a \ll 1$ (i.e. $U_0 \ll \hbar^2/ma^2$) we have

$$f = \tfrac{1}{3}a(\varkappa a)^2. \tag{2}$$

Formula (1) becomes inapplicable if U_0 and a are such that $\varkappa a$ is close to an odd multiple of $\frac{1}{2}\pi$. For such values of $\varkappa a$ the discrete spectrum of negative energy evels in the potential well includes one which is close to zero†, and the scattering s described by formulae which we shall derive in §66.

Problem 2. The same as Problem 1, but for scattering by a "potential hump": $U(r) = U_0$ for $r < a$, $U = 0$ for $r > a$.

Solution. This case is obtained from that of a potential well by changing the sign of U_0, i.e. $\varkappa \to i\varkappa$. Then, from (1),

$$f = \frac{\tanh \varkappa a - \varkappa a}{\varkappa a}$$

with $\varkappa = \sqrt{(2mU_0)}/\hbar$ as before). In particular, in the limit $\varkappa a \gg 1$ (i.e. for large U_0),

$$f = -a, \qquad \sigma = 4\pi a^2.$$

his corresponds to scattering by an impenetrable sphere of radius a; we notet at classical mechanics would give a result four times smaller ($\sigma = \pi a^2$).

---

† See §30, Problem 1. Equation (1) in the solution to that problem shows tha the energy level $|E| \ll U_0$ if $\sin [a\sqrt{(2mU_0)}/\hbar] \approx \pm 1$.

§66. Resonance scattering at low energies

Particular consideration must be given to the scattering of slow particles ($ka \ll 1$) in an attractive field when the discrete spectrum of negative energy levels includes an *s* state whose energy is small compared with the value of the field U within its radius of action a. We denote this level by $-\varepsilon\,(\varepsilon > 0)$. The energy E of the particle undergoing scattering, being small, is close to $-\varepsilon$, i.e. it is, as we say, almost in *resonance* with the level. This leads, as we shall see, to a considerable increase in the scattering cross-section.

The existence of the shallow level can be taken into account in scattering theory by means of a formal method based on the following arguments.

As in §65, we consider Schrödinger's equation in different parts of the field. The exact equation, with $\chi = r\psi$, is

$$\chi'' + (2m/\hbar^2)\,[E - U(r)]\,\chi = 0.$$

In the "inner" region $r \lesssim a$ we can neglect $(2mE/\hbar^2)\chi = k^2\chi$ in comparison with χ'':

$$\chi'' - (2m/\hbar^2)\,U(r)\chi = 0, \qquad r \sim a. \tag{66.1}$$

In the "outer" region ($r \gg a$), on the other hand, we can neglect U:

$$\chi'' + (2m/\hbar^2)\,E\chi = 0, \qquad r \gg a. \tag{66.2}$$

The solution of equation (66.2) must be "joined" at some r_1 (such that $1/k \gg r_1 \gg a$) to the solution of equation (66.1) which satisfies the boundary condition $\chi(0) = 0$; the joining condition is that the ratio χ'/χ should be continuous. This ratio does not depend on the normalisation factor in the wave function.

However, instead of considering the motion in the region $r \sim a$, we apply to the solution in the outer region a suitably chosen boundary condition on χ'/χ for small r; since the solution in the outer region varies only slowly as $r \to 0$, we can formally apply this condition at the point $r = 0$. The equation (66.1) for the region $r \sim a$ does not contain E; the boundary condition which replaces it must therefore also be independent of the energy of the particle. In other words, it must be of the form

$$[\chi'/\chi]_{r\to 0} = -\varkappa, \tag{66.3}$$

where $\varkappa$ is some constant. But, $\varkappa$ being independent of E, the same condition (66.3) must also apply to the solution of Schrödinger's equation for small negative energy $E = -\varepsilon$, i.e. to the wave function of the corresponding stationary state of the particle. For $E = -\varepsilon$ we have from (66.2)

$$\chi = \text{constant}\times e^{-r\sqrt{(2m\varepsilon)}/\hbar},$$

and substitution of this function in (66.3) shows that $\varkappa$ is a positive quantity,

$$\varkappa = \sqrt{(2m\varepsilon)}/\hbar. \tag{66.4}$$

Let us now apply the boundary condition (66.3) to the wave function for free motion,

$$\chi = \text{constant}\times\sin(kr+\delta_0),$$

which is the exact general solution of equation (66.2) for $E > 0$. Thus we have for the required phase δ_0

$$\begin{aligned}\cot\delta_0 &= -\varkappa/k\\ &= -\sqrt{(\varepsilon/E)}.\end{aligned} \tag{66.5}$$

Since the energy E is here restricted only by the condition $ka \ll 1$, and need not be small compared with ε, the phase δ_0 may not be small, and the same is true of the *s*-wave scattering amplitude.

The partial scattering amplitudes with $l \neq 0$ are again small. Hence we can again regard the total amplitude as being the same as the *s*-wave scattering amplitude:

$$\begin{aligned}f &= \frac{1}{2ik}(e^{2i\delta_0}-1)\\ &= 1/k(\cot\delta_0 - i),\end{aligned}$$

and substitution from (66.5) gives

$$f = -1/(\varkappa+ik). \tag{66.6}$$

This has a pole at $k = i\varkappa$, in accordance with the general result (§64). The total cross-section $\sigma = 4\pi\,|f|^2$ is

$$\sigma = \frac{4\pi}{\varkappa^2+k^2} = \frac{2\pi\hbar^2}{m}\frac{1}{E+\varepsilon}. \qquad (66.7)$$

Thus the scattering is again isotropic (the amplitude (66.6) does not depend on the direction), but the scattering cross-section varies with the energy, and in the resonance region ($E \sim \varepsilon$) it is large compared with the squared radius of action of the field a^2 (since $1/k \gg a$). The form of (66.7) does not depend on the detailed nature of the interaction of the particles at short distances, and is entirely determined by the energy of the resonance level.[†]

The above formula is somewhat more general than the assumption made in its derivation. Let the function $U(r)$ be slightly modified; this alters also the value of the constant $\varkappa$ in the boundary condition (66.3). By an appropriate change in $U(r)$, $\varkappa$ can be made to vanish, and then to become small and negative. This gives the same formulae (66.6) for the scattering amplitude and (66.7) for the cross-section. In the latter, however, the quantity $\varepsilon = \hbar^2\varkappa^2/2m$ is now simply a constant characteristic of the field $U(r)$, and not an energy level in that field. In such cases the field is said to have a *virtual level*, since, although there is no actual level close to zero, a slight change in the field would be sufficient to cause one to appear.[‡]

§67. **Born's formula**

The scattering cross-section can be calculated in a general form in a very important case, namely that where the scattering field may be regarded (in its action on the motion of the particle undergoing scat-

---

[†] Formula (66.7) was first derived by E. Wigner (1933); the idea of the derivation given here is due to H. A. Bethe and R. E. Peierls (1935).

[‡] As an example, we may mention that both types of resonance (at real and virtual levels) occur in the scattering of neutrons by protons. For the interaction of a neutron with a proton when their spins are parallel, there is a real level at $\varepsilon = 2.23$ Mev (the ground state of the deuteron). The interaction of a neutron with a proton when their spins are antiparallel corresponds to a virtual level at $\varepsilon = 0.067$ MeV.

tering) as a weak perturbation. We shall consider at the end of this section the conditions for the corresponding approximation to be applicable in the theory of scattering.

The unperturbed motion of a particle incident on a scattering centre with momentum $\mathbf{p} = \hbar\mathbf{k}$ is described by a plane wave $\psi^{(0)} = e^{i\mathbf{k}\cdot\mathbf{r}}$, which satisfies Schrödinger's equation

$$\triangle\psi^{(0)}+k^2\psi^{(0)} = 0.$$

We shall seek the solution of the exact equation

$$\triangle\psi+\left(k^2-\frac{2m}{\hbar^2}U\right)\psi = 0$$

in the form $\psi = \psi^{(0)}+\psi^{(1)}$; the small correction $\psi^{(1)}$ describing the scattered wave must satisfy the inhomogeneous (in $\psi^{(1)}$) equation

$$\triangle\psi^{(1)}+k^2\psi^{(1)} = \frac{2m}{\hbar^2}\psi^{(0)} = \frac{2m}{\hbar^2}e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{67.1}$$

in which the second-order term $\sim\psi^{(1)}U$ has been omitted.

The solution of this equation can be obtained directly by analogy with the equation of retarded potentials in electrodynamics:

$$\triangle\phi-\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = -4\pi\varrho,$$

where ϱ is a function of coordinates and time (see *Mechanics and Electrodynamics*, §79). Its solution is

$$\phi(\mathbf{r}, t) = \int\frac{1}{R}\varrho\left(\mathbf{r}', t-\frac{R}{c}\right)\mathrm{d}V', \quad \mathrm{d}V' = \mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z',$$

where $\mathbf{R} = \mathbf{r}-\mathbf{r}'$ is the radius vector from the volume element $\mathrm{d}V'$ to the "field point" r at which the value of ϕ is sought. If the time dependence of the function ϱ is given by a factor e^{-ikct}, then, with

$$\varrho = \varrho_0(\mathbf{r})\,e^{-ikct}, \quad \phi = \phi_0(\mathbf{r})\,e^{-ikct},$$

we have for ϕ_0 the equation

$$\triangle\phi_0+k^2\phi_0 = -4\pi\varrho_0, \tag{67.2}$$

and the solution is

$$\phi_0(\mathbf{r}) = \int \varrho_0(\mathbf{r}')\, e^{ikR}\, \mathrm{d}V'/R. \tag{67.3}$$

In view of the obvious analogy between equations (67.2) and (67.1), the solution of the latter equation can be written

$$\psi^{(1)}(\mathbf{r}) = -\frac{m}{2\pi\hbar^2} \int U(\mathbf{r}')\, e^{i(\mathbf{k}\cdot\mathbf{r}'+kR)} \frac{\mathrm{d}V'}{R}. \tag{67.4}$$

We can now easily derive an asymptotic expression for this function at large distances r from the scattering centre. When $r \gg r'$, $R = |\mathbf{r}-\mathbf{r}'| \approx r - \mathbf{r}'.\mathbf{n}'$, where $\mathbf{n}'$ is a unit vector in the direction of $\mathbf{k}'$; in the factor $1/R$ in the integrand of (67.4), it is sufficient to put simply $R \approx r$. Then

$$\psi^{(1)} = -\frac{m}{2\pi\hbar^2} \frac{e^{ikr}}{r} \int U(\mathbf{r}')\, e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}\, \mathrm{d}V',$$

where $\mathbf{k}' = k\mathbf{n}'$ is the wave vector of the particle after scattering. According to the definition (62.3), the coefficient of e^{ikr}/r in this function gives the required scattering amplitude; omitting the prime from the variables of integration, we can write the result as

$$f = -\frac{m}{2\pi\hbar^2} \int U(\mathbf{r})\, e^{-i\mathbf{q}\cdot\mathbf{r}}\, \mathrm{d}V. \tag{67.5}$$

Here we have introduced the vector

$$\mathbf{q} = \mathbf{k}'-\mathbf{k}, \tag{67.6}$$

whose absolute magnitude is

$$q = 2k \sin \tfrac{1}{2}\theta, \tag{67.7}$$

θ being the angle between $\mathbf{k}$ and $\mathbf{k}'$, i.e. the scattering angle. We see that the scattering amplitude with a particle momentum change $\hbar\mathbf{q}$ is determined by the corresponding Fourier component of the field $U(\mathbf{r})$. The cross-section for scattering into the solid angle element $\mathrm{d}o'$ is

$$\mathrm{d}\sigma = \frac{m^2}{4\pi^2\hbar^4} \left| \int U e^{-i\mathbf{q}\cdot\mathbf{r}}\, \mathrm{d}V \right|^2 \mathrm{d}o'. \tag{67.8}$$

This formula was first obtained by M. Born (1926); the corresponding approximation in collision theory is called the *Born approximation.*

Formula (67.8) can also be obtained by another method, starting from the general formula (35.6) of perturbation theory, which gives the transition probability between two states of the continuous spectrum. In the present case, we are concerned with a transition between states of a freely moving particle, with momenta **p** and **p**′, and the perturbation operator is represented by the function $U(\mathbf{r})$. As the "interval" of states $\mathrm{d}\nu_f$ we take the volume element $\mathrm{d}p_x' \,\mathrm{d}p_y' \,\mathrm{d}p_z'$ in momentum space. Then formula (35.6) becomes

$$\mathrm{d}w = \frac{2\pi}{\hbar} |U_{\mathbf{p}'\mathbf{p}}|^2 \delta\left(\frac{p'^2}{2m} - \frac{p^2}{2m}\right) \mathrm{d}p_x' \,\mathrm{d}p_y' \,\mathrm{d}p_z'. \tag{67.9}$$

The wave function of the final state must be normalised by the δ-function in momentum space (cf. the remark preceding (35.1)); according to (12.10), the plane wave thus normalised is

$$\psi_{\mathbf{p}'} = \frac{1}{(2\pi\hbar)^{3/2}} e^{i\mathbf{p}'\cdot\mathbf{r}/\hbar}. \tag{67.10}$$

The wave function of the initial state is normalised to unit current density:

$$\psi_{\mathbf{p}} = \sqrt{(m/p)}\, e^{i\mathbf{p}\cdot\mathbf{r}/\hbar}; \tag{67.11}$$

(cf. (21.6)). Then the "probability" (67.9) will have the dimensions of area, and is the differential scattering cross-section.

The δ-function appearing as a factor in (67.9) expresses the conservation of energy in elastic scattering, whereby the magnitude of the momentum is unchanged: $p' = p$. This δ-function can be eliminated by changing to "spherical polar coordinates" in momentum space (i.e. by replacing $\mathrm{d}p_x' \,\mathrm{d}p_y' \,\mathrm{d}p_z'$ by $p'^2 \,\mathrm{d}p' \,\mathrm{d}o' = \frac{1}{2} p' \,\mathrm{d}(p'^2) \,\mathrm{d}o'$) and integrating over p'^2. The integration amounts to replacing p' by p (and multiplying the whole expression by $2m$), and we obtain

$$\mathrm{d}\sigma = (2\pi m p/\hbar) \left|\int \psi_{\mathbf{p}'}^* U \psi_{\mathbf{p}} \,\mathrm{d}V\right|^2 \mathrm{d}o'. \tag{67.12}$$

Substituting in this expression (67.10) and (67.11), we return to (67.8). This derivation leads directly to the scattering cross-section, but leaves an undetermined phase in the scattering amplitude.

In formulae (67.5) and (67.8), the scattering field $U(\mathbf{r})$ is not assumed to be centrally symmetric. If, however, $U = U(r)$, the integration can be carried somewhat farther in a general form. To do so, we use spherical polar space coordinates r, ϑ, ϕ, with the polar axis in the direction of the vector $\mathbf{q}$; the polar angle is denoted by ϑ to distinguish it from the scattering angle θ. Then

$$\int U(r) e^{-i\mathbf{q}\cdot\mathbf{r}}\, dV = \int_0^\infty \int_0^{2\pi} \int_0^\pi U(r)\, e^{-iqr\cos\vartheta}\, r^2 \sin\vartheta\, d\vartheta\, d\phi\, dr.$$

The integration over ϑ and ϕ can be effected, and we thus obtain the following formula for the scattering amplitude in a centrally symmetric field:

$$f = -\frac{2m}{\hbar^2 q} \int_0^\infty U(r) \sin qr \cdot r\, dr. \qquad (67.13)$$

Let a be the radius of action of the field, and let us consider formula (67.13) in the limiting cases of small and large values of the product ka.

When $ka \ll 1$ (small velocities), we can put $\sin qr \approx qr$, so that the scattering amplitude is

$$f = -\frac{2m}{\hbar^2} \int_0^\infty U(r)\, r^2\, dr. \qquad (67.14)$$

Here the scattering is isotropic and independent of the velocity, in accordance with the general results of §65.

In the opposite limiting case of high velocities, when $ka \gg 1$, the scattering is markedly anisotropic and is directed forward in a narrow cone of angle $\Delta\theta \sim 1/ka$: since outside this cone the quantity q is large ($q \gg 1/a$), the factor $\sin qr$ is a rapidly oscillating function with variable sign in the region of action of the field ($r \lesssim a$), and the integral of its product with the slowly varying function U is almost zero.

Let us now ascertain the conditions for this approximation to be valid. The derivation of formula (67.5) was based on the approximate solution of Schrödinger's equation in the form $\psi = \psi^{(0)}+\psi^{(1)}$, with the assumption that $\psi^{(1)} \ll \psi^{(0)}$. It is sufficient to require the fulfilment of this condition in the most "dangerous" region near the scattering centre ($r = 0$); since $|\psi^{(0)}| = 1$, this means that $\psi^{(1)} \ll 1$. When $r = 0$, in the integral (67.4) we have $R = r'$, and so

$$\psi^{(1)}(0) = -\frac{m}{2\pi\hbar^2}\int U(\mathbf{r}')\, e^{i(\mathbf{k}\cdot\mathbf{r}'+kr')}\frac{\mathrm{d}V'}{r'}. \tag{67.15}$$

We can obtain approximate estimates of this integral for small and large values of ka.

When $ka \ll 1$, the exponential factor in the integrand can be replaced by unity, and then

$$\psi^{(1)}(0) \sim \frac{m\,|U|}{\hbar^2 a}a^3,$$

where $|U|$ is the order of magnitude of the field within its radius of action. Thus we have the condition

$$|U| \ll \hbar^2/ma^2, \quad ka \ll 1. \tag{67.16}$$

To estimate the integral for $ka \gg 1$, we first carry out the integration over the directions of $\mathbf{r}'$ (assuming the field to be a central field). Similarly to the derivation of (67.13), we find

$$\begin{aligned}\psi^{(1)}(0) &= -\frac{m}{2\pi\hbar^2}\int_0^\infty\int_0^\pi U(r')\, e^{ikr'(\cos\vartheta+1)}.2\pi\sin\vartheta\,\mathrm{d}\vartheta.r'\,\mathrm{d}r' \\ &= -\frac{m}{\hbar^2 ik}\int_0^\infty U(r')\,(e^{2ikr'}-1)\,\mathrm{d}r'.\end{aligned}$$

When $ka \gg 1$, the integral of the term containing the oscillating factor $\exp(2ikr')$ is almost zero, and the integral of the second term is approximately $|U|\,a$. Thus we obtain the condition

$$|U| \ll \hbar^2 ka/ma^2 = \hbar v/a, \quad ka \gg 1. \tag{67.17}$$

It is evident that, if the field satisfies the condition (67.16), it also satisfies the weaker condition (67.17) when $ka \gg 1$. Thus in this case the Born approximation is valid for both small and large velocities. It is in any case valid for sufficiently large velocities, by (67.17), even if the condition (67.16) for it to be applicable at small velocities is not satisfied.

PROBLEMS

PROBLEM 1. Determine, in the Born approximation, the scattering cross-section for a spherical potential well: $U = -U_0$ for $r < a$, $U = 0$ for $r > a$.

SOLUTION. The calculation of the integral in (67.13) gives

$$d\sigma = 4a^2\left(\frac{mU_0a^2}{\hbar^2}\right)^2 \frac{(\sin qa - qa\cos qa)^2}{(qa)^6}\, do.$$

The integration over all angles (which is conveniently effected by using the variable $q = 2k\sin\frac{1}{2}\theta$ and replacing do by $2\pi q\, dq/k^2$) gives the total scattering cross-section

$$\sigma = \frac{2\pi}{k^2}\left(\frac{mU_0a^2}{\hbar^2}\right)^2\left[1 - \frac{1}{(2ka)^2} + \frac{\sin 4ka}{(2ka)^3} - \frac{\sin^2 2ka}{(2ka)^4}\right].$$

In the limiting cases this formula gives

$$\sigma = \frac{16\pi a^2}{9}\left(\frac{mU_0a^2}{\hbar^2}\right)^2 \quad \text{for} \quad ka \ll 1,$$

$$\sigma = \frac{2\pi}{k^2}\left(\frac{mU_0a^2}{\hbar^2}\right)^2 \quad \text{for} \quad ka \gg 1.$$

The first of these expressions corresponds to the scattering amplitude (2) in §65, Problem 1, found there by a different method.

PROBLEM 2. The same as Problem 1, but in a field $U = (\alpha/r)e^{-r/a}$.

SOLUTION. The calculation of the integral in (67.13) gives

$$d\sigma = 4a^2\left(\frac{\alpha ma}{\hbar^2}\right)^2 \frac{do}{(q^2a^2+1)^2}. \qquad (1)$$

The total cross-section is

$$\sigma = 16\pi a^2\left(\frac{\alpha ma}{\hbar^2}\right)^2 \frac{1}{4k^2a^2+1}.$$

The condition for these formulae to be applicable is found from (67.16) and (67.17) with α/a instead of U: $\alpha ma/\hbar^2 \ll 1$ or $\alpha/\hbar v \ll 1$.

This potential represents a "screened" Coulomb field with a screening radius a. When $a \to \infty$, we have a pure Coulomb field, and the differential cross-section (1) becomes Rutherford's formula (§68).

§68. Rutherford's formula

Let us now apply Born's formula to scattering in a Coulomb field, taking the particular case of scattering of particles with charge e by nuclei with charge Ze; then $U = Ze^2/r$.

According to (67.5), the problem amounts to calculating the Fourier component of the function $1/r$. Instead of a direct calculation, it is more convenient to start from the differential equation

$$\triangle(1/r) = -4\pi\delta(\mathbf{r}), \tag{68.1}$$

satisfied by the function $1/r$ (see *Mechanics and Electrodynamics*, (59.10)).[†] But, with a view to certain other applications, we shall first consider the more general case of a function $\phi(\mathbf{r})$ that satisfies the equation

$$\triangle\phi = -4\pi\varrho(\mathbf{r}) \tag{68.2}$$

with a given function $4\pi\varrho(\mathbf{r})$ on the right-hand side.

We expand the function $\phi(\mathbf{r})$ as a Fourier integral:

$$\phi(\mathbf{r}) = \int e^{i\mathbf{q}\cdot\mathbf{r}}\phi_{\mathbf{q}}\,\mathrm{d}^3q/(2\pi)^3, \quad \mathrm{d}^3q = \mathrm{d}q_x\,\mathrm{d}q_y\,\mathrm{d}q_z. \tag{68.3}$$

Then

$$\phi_{\mathbf{q}} = \int \phi(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}\,\mathrm{d}V. \tag{68.4}$$

Taking the Laplacian of both sides of (68.3) and differentiating under the integral sign, we have

$$\triangle\phi = -\int q^2 e^{i\mathbf{q}\cdot\mathbf{r}}\phi_{\mathbf{q}}\,\mathrm{d}^3q/(2\pi)^3.$$

This means that the Fourier component of $\triangle\phi$ is $(\triangle\phi)_{\mathbf{q}} = -q^2\phi_{\mathbf{q}}$. We can also derive $(\triangle\phi)_{\mathbf{q}}$ by taking the Fourier component of each side of equation (68.2): $(\triangle\phi)_{\mathbf{q}} = -4\pi\varrho_{\mathbf{q}}$. Equating the two expressions, we have

$$\begin{aligned}\phi_{\mathbf{q}} &= (4\pi/q^2)\varrho_{\mathbf{q}}\\ &= (4\pi/q^2)\int \varrho(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}}\,\mathrm{d}V.\end{aligned} \tag{68.5}$$

[†] Another method of calculation is to begin with a "screened" Coulomb field, and then make the screening radius tend to infinity (see §67, Problem 2).

If the function $\phi = 1/r$, then $\varrho = \delta(\mathbf{r})$, and the integral on the right-hand side of (68.5) is unity, so that

$$(1/r)_{\mathbf{q}} = 4\pi/q^2. \tag{68.6}$$

The scattering amplitude in a Coulomb field is, by (67.5) and (67.7),

$$f(\theta) = -\frac{mZe^2}{2\pi\hbar^2}\cdot\frac{4\pi}{q^2} = -\frac{Ze^2}{2mv^2}\cdot\frac{1}{\sin^2\frac{1}{2}\theta}, \tag{68.7}$$

with the velocity v of the particles undergoing scattering: $\hbar k = mv$. Hence the scattering cross-section is

$$d\sigma = (Ze^2/2mv^2)^2\, do/\sin^4\tfrac{1}{2}\theta, \tag{68.8}$$

in accordance with the classical Rutherford's formula.

Because of the slow decrease of the Coulomb field, it is impossible to distinguish a finite region of space in which U is considerably greater than outside that region. The condition for the Born approximation to be applicable to scattering in this field is obtained from (67.17), the parameter a being replaced by the variable distance r:

$$Ze^2/\hbar v \ll 1. \tag{68.9}$$

The opposite inequality is given by (63.2) as the condition for quasi-classical scattering in a Coulomb field: $Ze^2/\hbar v \gg 1$. In this case, the scattering must certainly be in accordance with Rutherford's formula. Hence we see that Rutherford's formula is obtained in the limiting cases of both large and small velocities. This leads us to expect the result given by the quantum theory of scattering using the exact solution of Schrödinger's equation in a Coulomb field: the exact quantum-mechanical formula for the scattering cross-section is the same as the classical Rutherford's formula (N. F. Mott and W. Gordon 1928).†

---

† To avoid misunderstanding, however, it should be emphasised that the same does not apply to the formula (68.7) for the scattering amplitude: the exact expression for $f(\theta)$ differs from (68.7) by a phase factor depending on θ and v, which becomes unity only when the condition (68.9) is satisfied.

§69. Collisions of like particles

The case where two identical particles collide requires special consideration. The identity of the particles leads in quantum mechanics, as we know (§46), to the appearance of a peculiar exchange interaction between them. This has an important effect on scattering also (N. F. Mott 1930).

Let us consider the particular case of a collision between two like particles with spin $\frac{1}{2}$ (two electrons, or two nucleons). The orbital wave function of a system of two such particles must be symmetric with respect to interchange of the particles if the total spin of the system $S = 0$, and antisymmetric if $S = 1$ (§46). The wave function which describes the scattering, and which is obtained by solving the usual Schrödinger's equation, must therefore be symmetrised or antisymmetrised with respect to the particles. An interchange of the particles is equivalent to reversing the direction of the radius vector joining them. In the coordinate system in which the centre of mass is at rest, this means that r remains unchanged, while the angle θ is replaced by $\pi-\theta$ (and so $z = r\cos\theta$ becomes $-z$). Hence, instead of the asymptotic expression (62.3) for the wave function, we must write

$$\psi = e^{ikz} \pm e^{-ikz} + e^{ikr}[f(\theta) \pm f(\pi-\theta)]/r. \tag{69.1}$$

By virtue of the identity of the particles it is, of course, impossible to say which of them scatters and which is scattered. In the coordinate system in which the centre of mass is at rest, we have two equal incident plane waves, propagated in opposite directions (e^{ikz} and e^{-ikz} in (69.1)). The outgoing spherical wave in (69.1) takes into account the scattering of both particles and, the probability current calculated from it gives the probability that either of the particles will be scattered into the element do of solid angle considered. The scattering cross-section is the ratio of this current to the current density in either of the incident plane waves, i.e. is given, as before, by the squared modulus of the coefficient of e^{ikr}/r in the wave function (69.1).

Thus, if the total spin of the colliding particles is zero, the scattering cross-section is of the form

$$d\sigma_0 = |f(\theta)+f(\pi-\theta)|^2\, do, \tag{69.2}$$

while if the total spin is unity, it is

$$d\sigma_1 = |f(\theta)-f(\pi-\theta)|^2\, do. \tag{69.3}$$

The appearance of the "interference" term $f(\theta)f^*(\pi-\theta)+f^*(\theta)f(\pi-\theta)$ characterises the exchange interaction. If the particles were different, as they are in classical mechanics, the probability that either of them would be scattered into a given element of solid angle do would simply be equal to the sum of the probabilities that one particle is deviated through an angle θ and the other through $\pi-\theta$; in other words, the cross-section would be

$$\{|f(\theta)|^2+|f(\pi-\theta)|^2\}\, do. \tag{69.4}$$

In formulae (69.2), (69.3) it is supposed that the total spin of the colliding particles has a definite value. If the system is not in a definite spin state, then to determine the cross-section it is necessary to average over all possible spin states, assuming them to be all equally probable. Of the total number of $2\times 2 = 4$ different spin states of a system of two particles with spin $\frac{1}{2}$, one state corresponds to zero total spin $\left(\text{spin projections } \frac{1}{2}, -\frac{1}{2}\right)$ and three states to unit total spin $\left(\text{spin projections } \frac{1}{2}, \frac{1}{2};\ -\frac{1}{2}, -\frac{1}{2};\ -\frac{1}{2}, \frac{1}{2}\right)$. Then the probability that the system will have $S = 0$ and $S = 1$ is $\frac{1}{4}$ and $\frac{3}{4}$ respectively. Hence the cross-section is

$$\begin{aligned} d\sigma &= \tfrac{1}{4}\, d\sigma_0+\tfrac{3}{4}\, d\sigma_1 \\ &= \{|f(\theta)|^2+|f(\pi-\theta)|^2-\tfrac{1}{2}[f(\theta)f^*(\pi-\theta)+f^*(\theta)f(\pi-\theta)]\}\, do. \end{aligned} \tag{69.5}$$

As an example, we shall consider the collision of two fast electrons interacting by Coulomb's law ($U = e^2/r$). If the condition (68.9) $e^2/\hbar v \ll 1$ is satisfied (where v is the velocity of the relative motion of the particles), we can use for the amplitude the Born-approximation expression (68.7). Here it must be noted that m denotes the reduced mass of the two particles, which in this case is $\frac{1}{2}m_e$ if m_e is the electron mass. Substitution of (68.7) in (69.5) gives

$$d\sigma = \left(\frac{e^2}{m_e v^2}\right)^2 \left[\frac{1}{\sin^4 \frac{1}{2}\theta} + \frac{1}{\cos^4 \frac{1}{2}\theta} - \frac{1}{\sin^2 \frac{1}{2}\theta \cos^2 \frac{1}{2}\theta}\right] do. \quad (69.6)$$

This refers to the system in which the centre of mass of the two electrons is at rest. For the laboratory system, where one of the electrons is at rest before the collision, the corresponding formula is obtained by simply replacing θ by 2ϑ (see (62.2)). Then

$$d\sigma = \left(\frac{2e^2}{m_e v^2}\right)^2 \left[\frac{1}{\sin^4 \vartheta} + \frac{1}{\cos^4 \vartheta} - \frac{1}{\sin^2 \vartheta \cos^2 \vartheta}\right] \cos \vartheta \, do, \quad (69.7)$$

where do is the element of solid angle in the new system of coordinates; when θ is replaced by 2ϑ, do must be replaced by 4 cos ϑ do, since $\sin \theta \, d\theta \, d\phi = 4 \cos \vartheta \sin \vartheta \, d\vartheta \, d\phi$. The final terms in (69.6) and (69.7) constitute the difference from the classical formulae (see *Mechanics and Electrodynamics*, §16).

PROBLEM

Determine the scattering cross-section for two identical particles with spin $\frac{1}{2}$, polarised in directions which differ by an angle α.

SOLUTION. The dependence of the cross-section σ on the polarisations of the particles must be expressed by a term proportional to the scalar $\bar{\mathbf{s}}_1 . \bar{\mathbf{s}}_2$, the product of the mean values of the spin vectors of the two particles; for particles polarised in directions making an angle α with each other, this product is $\frac{1}{4} \cos \alpha$. We look for σ in the form $a + 4b\bar{\mathbf{s}}_1 . \bar{\mathbf{s}}_2$. When the particles are unpolarised, the second term disappears ($\bar{\mathbf{s}}_1 = \bar{\mathbf{s}}_2 = 0$) and according to (69.5) $\sigma = a = \frac{1}{4}(\sigma_0 + 3\sigma_1)$. If the two particles are polarised in the same direction ($\alpha = 0$), i.e. have equal spin projections on the same direction, the total spin of the system is certainly $S = 1$; hence $\sigma = a + b = \sigma_1$. Determining a and b from these two equations, we find

$$\sigma = \tfrac{1}{4}\{(\sigma_0 + 3\sigma_1) + (\sigma_1 - \sigma_0) \cos \alpha\}.$$

§70. Elastic collisions between fast electrons and atoms

Elastic collisions between fast electrons and atoms can be treated by means of the Born approximation if the velocity of the incident electron is large compared with those of the atomic electrons.

Owing to the large difference in mass between the electron and the atom, the latter may be regarded as at rest during the collision, and the system of coordinates in which the centre of mass is fixed is the same as the laboratory system, in which the atom is fixed. Then $\mathbf{p}$ and $\mathbf{p}'$ in the formulae of §67 denote the momenta of the electron before and after the collision, m the mass of the electron, and the angle θ is the same as the angle of deviation ϑ of the electron.

In §67 we have calculated the matrix element $U_{\mathbf{p}'\mathbf{p}}$ of the interaction energy with respect to the wave functions of a free particle before and after the collision. In a collision with an atom it is necessary to take into account also the wave functions describing the internal state of the atom. Hence $U_{\mathbf{p}'\mathbf{p}}$ in (67.8) must be replaced by the matrix element of the electron-atom interaction energy U, taken with respect to the wave functions of the electron and the atom. Since the state of the atom is unchanged by an elastic collision, the matrix element is diagonal with respect to this state. Thus the formula for the cross-section must be written

$$\mathrm{d}\sigma = \frac{m^2}{4\pi^2\hbar^4}\left|\int\int \psi_0^* U e^{-i\mathbf{q}\cdot\mathbf{r}}\psi_0 \,\mathrm{d}\tau\,\mathrm{d}V\right|^2 \mathrm{d}o, \qquad (70.1)$$

where ψ_0 is the atomic wave function (depending on the coordinates of all the Z electrons in the atom), and $\mathrm{d}\tau = \mathrm{d}V_1 \ldots \mathrm{d}V_Z$ is an element of the configuration space of the atomic electrons.

The integral

$$\int \psi_0^* U \psi_0 \,\mathrm{d}\tau$$

is the energy of interaction of the electron with the atom, averaged with respect to the state of the latter. It can also be written as $e\phi(r)$, where $\phi(r)$ is the potential of the electric field due to the mean distribution of charges in the atom.

Denoting the density of this charge distribution by $\varrho(\mathbf{r})$, we have, for the potential ϕ, Poisson's equation:

$$\triangle\phi = -4\pi\varrho(\mathbf{r}).$$

The required matrix element in (70.1) is the Fourier component $e\phi_{\mathbf{q}}$. According to (68.5), this can be calculated by means of the Fourier component of the charge density ϱ. The latter consists of the electron and nuclear charges:

$$\varrho = -|e|\,n(r)+Z\,|e|\,\delta(\mathbf{r}),$$

where $n(r)$ is the electron number density in the atom. Multiplying by $e^{-i\mathbf{q}\cdot\mathbf{r}}$ and integrating, we have

$$\int \varrho e^{-i\mathbf{q}\cdot\mathbf{r}}\,\mathrm{d}V = -|e|\int n e^{-i\mathbf{q}\cdot\mathbf{r}}\,\mathrm{d}V+Z\,|e|.$$

Thus we obtain for the matrix element in question the expression

$$\iint \psi_0^* U e^{-i\mathbf{q}\cdot\mathbf{r}}\psi_0\,\mathrm{d}\tau\,\mathrm{d}V = -\frac{4\pi e^2}{q^2}[Z-F(q)], \qquad (70.2)$$

where $F(q)$ is defined by the formula

$$F(q) = \int n e^{-i\mathbf{q}\cdot\mathbf{r}}\,\mathrm{d}V \qquad (70.3)$$

and is called the *atomic form factor*. It is a function of the scattering angle and of the velocity of the incident electron.

Finally, substituting (70.2) in (70.1), we obtain the following expression for the cross-section for the elastic scattering of fast electrons by an atom†:

$$\mathrm{d}\sigma = \frac{4m^2e^4}{\hbar^4q^4}[Z-F(q)]^2\,\mathrm{d}o. \qquad (70.4)$$

The variable $\hbar q$ is the momentum transfer from the electron to the atom. It is related to the electron velocity v and the scattering angle ϑ by

$$q = (2mv/\hbar)\sin\tfrac{1}{2}\vartheta; \qquad (70.5)$$

cf. (67.7).

Let us consider the limiting case of small q (small compared with $1/a$, where a is of the order of magnitude of the dimensions of the atom

---

† We are neglecting exchange effects between the fast electron which undergoes scattering and the atomic electrons, i.e. we do not symmetrise the wave function of the system. The legitimacy of this procedure follows because the interference terms in the cross-section are removed by the rapid oscillation of the wave function of the incident electron within the volume of the atom over which the slowly varying wave function of the atomic electrons extends.

($qa \ll 1$)). Small scattering angles correspond to small $q : \vartheta \ll v_0/v$, where $v_0 \sim \hbar/ma$ is of the order of magnitude of the velocities of the atomic electrons.

Let us expand $F(q)$ as a series of powers of q. The zero-order term is $\int n \,\mathrm{d}V$, which is the total number Z of electrons in the atom. The first-order term is proportional to $\int \mathbf{r}n(r)\,\mathrm{d}V$, i.e. to the mean value of the dipole moment of the atom; this vanishes identically (see §54).

We must therefore continue the expansion up to the second-order term, obtaining

$$Z-F(q) = \tfrac{1}{2}\int (\mathbf{q}\cdot\mathbf{r})^2 n\,\mathrm{d}V = \tfrac{1}{6}q^2\int nr^2\,\mathrm{d}V;$$

substituting in (70.4), we obtain

$$\mathrm{d}\sigma = \left|\frac{me^2}{3\hbar^2}\int nr^2\,\mathrm{d}V\right|^2 \mathrm{d}o. \tag{70.6}$$

Thus, in the range of small angles, the cross-section is independent of the scattering angle, and is given by the mean square distance of the atomic electrons from the nucleus.

In the opposite limiting case of large q ($qa \gg 1$), the factor $e^{-i\mathbf{q}\cdot\mathbf{r}}$ in the integrand in (70.3) is a rapidly oscillating function, and therefore the whole integral is nearly zero. Consequently, we can neglect $F(q)$ in comparison with Z, so that

$$\mathrm{d}\sigma = \left(\frac{Ze^2}{2mv^2}\right)^2 \frac{\mathrm{d}o}{\sin^4\frac{1}{2}\vartheta}. \tag{70.7}$$

In other words, we have Rutherford scattering at the nucleus of the atom.

PROBLEM

Calculate the cross-section for the elastic scattering of fast electrons by a hydrogen atom in the ground state.

SOLUTION. The wave function of the normal state of the hydrogen atom is (in ordinary units) $\psi = \pi^{-1/2}e^{-r/a_B}$, where $a_B = \hbar^2/me^2$ is the Bohr radius (see (31.15)). The electron density is $n = |\psi|^2$. The integration over angles in (70.3) is effected as in the derivation of formula (67.13), and gives

$$F = \frac{4\pi}{q}\int\limits_0^\infty n(r)\sin qr . r\,\mathrm{d}r = (1+\tfrac{1}{4}a_B q^2)^{-2}.$$

Substituting in (70.4), we obtain

$$\mathrm{d}\sigma = 4a_B^2\frac{(8+a_B^2q^2)^2}{(4+a_B^2q^2)^4}\,\mathrm{d}o.$$

The total cross-section is conveniently calculated by putting

$$\mathrm{d}o = 2\pi\sin\vartheta\,\mathrm{d}\vartheta = 2\pi(\hbar/mv)^2\,q\,\mathrm{d}q$$

and integrating over q; here, of course, only the term of the lowest degree in $1/v$ need be retained (in the Born approximation). The result is

$$\sigma = (7\pi/3)\,(\hbar/mv)^2.$$

CHAPTER 10

INELASTIC COLLISIONS

§71. The principle of detailed balancing

Collisions are said to be *inelastic* when they are accompanied by a change in the internal state of the colliding particles. Here we understand "a change in the internal state" in the widest sense; in particular, the very nature of the particles or their number may be altered. For example, the change may consist in the excitation or ionisation of atoms, the excitation or disintegration of nuclei, and so on. Where a collision (e.g. a nuclear reaction) may be accompanied by various physical processes, these are referred to as various *channels* of the reaction.

Using the symmetry of the theory under time reversal, we can establish a very general relation between the probabilities or cross-sections for various inelastic processes. Let us consider, in particular, reactions of the form $a+b \rightarrow c+d$, with two particles in the initial state and two in the final state.

For convenience, we shall first of all suppose that the particles are moving in some large but finite volume Ω (and later take the limit $\Omega \rightarrow \infty$). Then the spectrum of free motion of the particles is not continuous, but discrete with the energy levels very close together, the intervals tending to zero as $\Omega \rightarrow \infty$ (cf. the end of §27).

Let w_{fi} be the probability of a transition of the system of colliding particles from a state i to a state f.[†] Each of these states is characterised

---

[†] The suffix for the final state is written on the left of that for the initial state, for uniformity with the customary order of suffixes in the matrix elements of transitions.

by the types of particle and by definite velocity vectors and spin projections of these particles.[†] Time reversal changes, firstly, the signs of the velocities and spin projections[‡]; the states obtained from i and f by these changes will be denoted by i^* and f^* and called the *time-reversed* states corresponding to i and f. Moreover, the initial state becomes the final state, and vice versa. From the symmetry of the equations of quantum mechanics under time reversal, the probabilities of the transitions $i \to f$ and $f^* \to i^*$ must be equal:

$$w_{fi} \to w_{i^*f^*}. \qquad (71.1)$$

This equation represents the *principle of detailed balancing.*

Let us now change from probabilities to reaction cross-sections. We denote by $\mathbf{p}_i$, $\mathbf{v}_i$ and $\mathbf{p}_f$, $\mathbf{v}_f$ the momenta and velocities of the relative motion of the two initial and two final particles. Let $d\sigma_{fi}$ be the cross-section for collisions such that $\mathbf{v}_f$ is in the direction of the solid-angle element do_f (in the centre-of-mass system of the two particles). The total energy of the two particles is, of course, the same before and after the collision ($E_i = E_f$), but we shall define a cross-section that is formally related to a range dE_f of values of the energy in the final state, regarded as a variable quantity. This cross-section must be written in the form

$$d\sigma_{fi} . \delta(E_f - E_i)\, dE_f. \qquad (71.2)$$

The δ-function here ensures that the law of conservation of energy is satisfied.

The collision cross-section is, by definition, obtained by dividing the probability of the relevant process by the incident particle current density. The latter is v_i/Ω (the factor $1/\Omega$ is the particle number density corresponding to one particle in the volume Ω). It must also be noted

---

[†] For "complex" particles (atoms and nuclei), the "spin" must here be taken as the total intrinsic angular momentum, made up of the spins and the orbital angular momenta of the internal motion of the constituent electrons and nucleons.

[‡] Each physical quantity behaves under time reversal in a definite manner that does not, of course, depend on the validity of a particular mechanics. The behaviour of the angular momentum is evident from the classical formula $\mathbf{r} \times \mathbf{p} = m\mathbf{r} \times \mathbf{v}$: it changes sign together with the velocity.

that the cross-section (71.2) relates to the ranges $\mathrm{d}o_f$ and $\mathrm{d}E_f$, whereas the probability w_i relates to specific values of v_f and E_f. Hence, to find the cross-section $\mathrm{d}\sigma_{fi}$, we must multiply w_{fi} by the number of quantum states belonging to the specified ranges of directions and of values of the velocity $\mathbf{v}_f$ or the momentum $\mathbf{p}_f$. This number is

$$\Omega p_f^2\,\mathrm{d}p_f\,\mathrm{d}o_f/(2\pi\hbar)^3;$$

cf. (27.8).

Summarizing these arguments, we can derive the following relation between the cross-section and the probability:

$$\mathrm{d}\sigma_{fi}\,.\,\delta(E_f-E_i)\,\mathrm{d}E_f = \frac{w_{fi}}{v_i/\Omega}\,\frac{\Omega p_f^2\,\mathrm{d}p_f\,\mathrm{d}o_f}{(2\pi\hbar)^3}\,.$$

Hence

$$w_{fi} = \frac{(2\pi\hbar)^3}{\Omega^2}\,\frac{v_i\,\mathrm{d}\sigma_{fi}\delta(E_f-E_i)\,\mathrm{d}E_f}{p_f^2\,\mathrm{d}p_f\,\mathrm{d}o_f} = \delta(E_f-E_i)\,\frac{(2\pi\hbar)^3\,v_i v_f}{\Omega^2}\,\frac{\mathrm{d}\sigma_{fi}}{p_f^2\,\mathrm{d}o_f};$$

here the velocity v_f is written for $\mathrm{d}E_f/\mathrm{d}p_f$, the equality of these being obvious from the fact that the kinetic energy of the relative motion is one term in E_f. Lastly, writing the probability $w_{f^*i^*}$ in a similar form, equating the two expressions, and cancelling the common factors, we get

$$\frac{\mathrm{d}\sigma_{fi}}{p_f^2\,\mathrm{d}o_f} = \frac{\mathrm{d}\sigma_{i^*f^*}}{p_i^2\,\mathrm{d}o_i}\,. \tag{71.3}$$

This relation gives the principle of detailed balancing in terms of cross-sections. The volume Ω does not appear in it, and the form of the relation is therefore unchanged in the limit $\Omega \to \infty$.

The equations (71.1) and (71.3) relate the probabilities or cross-sections for the two processes $i \to f$ and $f^* \to i^*$, which, though not strictly opposite processes (as $i \to f$ and $f \to i$ would be), are physically very close to being opposite.

The difference between the transitions $i \to f$ and $i^* \to f^*$ disappears altogether if we consider the cross-sections integrated over all directions of $\mathbf{p}_f$, summed over directions of the spins $\mathbf{s}_{1f}$, $\mathbf{s}_{2f}$ of the final particles, and averaged over the directions of the momentum $\mathbf{p}_i$ and spins $\mathbf{s}_{1i}$, $\mathbf{s}_{2i}$ of the initial particles. Let this cross-section be denoted by $\overline{\sigma_{fi}}$:

$$\overline{\sigma_{fi}} = \frac{1}{4\pi(2s_{1i}+1)(2s_{2i}+1)} \sum_{(m_s)} \iint \mathrm{d}\sigma_{fi}\, \mathrm{d}o_i; \tag{71.4}$$

the sum is taken over the spin projections of the particles; the factor before the summation and integration signs arises because we average, not sum, over quantities pertaining to the initial particles. Writing (71.3) in the form

$$p_i^2\, \mathrm{d}\sigma_{fi}\, \mathrm{d}o_i = p_f^2\, \mathrm{d}\sigma_{i^*f^*}\, \mathrm{d}o_f$$

and carrying out the operations mentioned, we get the required relation:

$$g_i p_i^2 \overline{\sigma_{fi}} = g_f p_f^2 \overline{\sigma_{if}}. \tag{71.5}$$

Here g_i and g_f denote the quantities

$$\left.\begin{aligned} g_i &= (2s_{1i}+1)(2s_{2i}+1), \\ g_f &= (2s_{1f}+1)(2s_{2f}+1), \end{aligned}\right\} \tag{71.6}$$

which give the number of possible spin orientations of the initial pair or the final pair of particles; they are called the *spin statistical weights* of the states *i* and *f*.

PROBLEMS[†]

Problem 1. Find the relation between the cross-sections σ_{pe} for the photoelectric effect (ionisation of an atom with absorption of a photon $\hbar\omega$) and σ_{rr} for radiative recombination (capture of a free electron by an ion to form a neutral atom with simultaneous emission of a photon).

[†] In these problems, we make use of certain concepts, relating to the photon, that will be explained in the next chapter.

Solution. The states i and f in (71.5) are in this case the states of the ion+ +electron and atom+photon systems. The required relation is

$$(2J_i+1)\,p^2\overline{\sigma_{rr}} = 2(2J_a+1)\,(\hbar\omega/c)^2\,\overline{\sigma_{pe}},$$

where J_i and J_a are the angular momenta of the ion and the atom, $p = mv$ is the momentum of the electron incident on the ion at rest, $\hbar\omega/c$ is the photon momentum, and the factor 2 is the statistical weight of the photon (two directions of polarisation).

Problem 2. Find the relation between the cross-sections for photodisintegration of a deuteron and radiative capture of a proton by a neutron.

Solution. The spin statistical weight of the neutron+proton system is $2\times2 = 4$; the statistical weight of the deuteron (in the ground state with $S = 1$)+ +photon system is $3\times2 = 6$. Hence $4p^2\overline{\sigma_{rc}} = 6(\hbar\omega/c)^2\overline{\sigma_{pd}}$, where p is the momentum of the relative motion of the colliding proton and neutron. This momentum is related to the deuteron binding energy I and the energy $\hbar\omega$ of the γ quantum emitted in radiative capture by the law of conservation of energy: $I+p^2/M = \hbar\omega$ (the reduced mass is $\frac{1}{2}M$, where M is the nucleon mass). The final result is

$$2Mc^2(\hbar\omega - I)\overline{\sigma_{rc}} = 3(\hbar\omega)^2\overline{\sigma_{pd}}\,.$$

§72. Elastic scattering in the presence of inelastic processes

The existence of inelastic channels has a certain effect on the properties of elastic scattering also.

The wave function ψ describing the process of elastic scattering consists of an incident plane wave and an outgoing spherical wave. It can also be represented as the sum of ingoing and outgoing waves, as in §62, these being "partial" waves (i.e. corresponding to definite values of the orbital angular momentum l). In (62.7), however, the amplitudes of each pair of ingoing and outgoing partial waves were equal, the coefficients of e^{-ikr} and e^{ikr} in the brackets being both of modulus unity. In purely elastic scattering this corresponds to the physical significance of the problem, but when there are inelastic channels the amplitudes of the outgoing waves must be less than those of the ingoing waves. The asymptotic expression for ψ will therefore be given by

$$\psi = \frac{i}{2kr}\sum_{l=0}^{\infty}(2l+1)\,P_l(\cos\theta)\,[(-1)^l e^{-ikr}-S_l e^{ikr}], \qquad (72.1)$$

the coefficients which replace the exp $(2i\delta_l)$ of (62.7) being complex quantities S_l, with moduli less than unity. The elastic scattering amplitude is given correspondingly by an expression different from (62.8):

$$f(\theta) = \frac{1}{2ik}\sum_{l=0}^{\infty}(2l+1)(S_l-1)P_l(\cos\theta). \tag{72.2}$$

For the total elastic scattering cross-section σ_e we have, instead of (62.9), the formula

$$\sigma_e = \sum_{l=0}^{\infty}\frac{\pi}{k^2}(2l+1)\,|1-S_l|^2. \tag{72.3}$$

The total inelastic scattering cross-section or *reaction cross-section* σ_r for all possible channels can also be expressed in terms of the S_l. To do so, we need only note that for each value of l the intensity of the outgoing wave is reduced in the ratio $|S_l|^2$ in comparison with that of the ingoing wave. This reduction must be ascribed entirely to inelastic scattering. It is therefore clear that

$$\sigma_r = \sum_{l=0}^{\infty}\frac{\pi}{k^2}(2l+1)(1-|S_l|^2), \tag{72.4}$$

and the total cross-section is

$$\sigma_t = \sigma_e+\sigma_r = \sum_{l=0}^{\infty}\frac{\pi}{k^2}(2l+1)(2-S_l-S_l^*). \tag{72.5}$$

Each term in the sums (72.3) and (72.4) is the partial cross-section for elastic or inelastic scattering of particles with orbital angular momentum l. The value $S_l = 1$ corresponds to the complete absence of scattering (with a given l). The case $S_l = 0$ corresponds to total "absorption" of particles with a given l (the outgoing partial wave with this l then does not appear in (72.1)); the cross-sections for elastic and inelastic scattering are equal. Although elastic scattering can occur without inelastic scattering (when $|S_l| = 1$), the opposite situation is impossible: the presence of inelastic scattering necessarily implies the simultaneous presence of elastic scattering.

When $\theta \to 0$, the elastic scattering amplitude (72.2) tends to the value

$$f(0) = \frac{1}{2k}\sum_{l=0}^{\infty}(2l+1)\,i(1-S_l).$$

Comparing this with (72.5), we find the following relation between the imaginary part of the amplitude of elastic scattering through zero angle and the total cross-section for scattering in all channels:

$$\operatorname{im} f(0) = k\sigma_t/4\pi.$$

This is called the *optical theorem* for scattering.

§73. Inelastic scattering of slow particles

The derivation of the limiting law of elastic scattering at low energies given in §65 can easily be generalised to the case where inelastic processes are involved.

As before, the s-wave scattering (with $l = 0$) is the most important at low energies. According to the results of §65, the quantity $S_0 = \exp(2i\delta_0)$ for small k is

$$S_0 \approx 1+2i\delta_0 = 1+2ik\beta, \tag{73.1}$$

where $\beta = c_2/c_1$ is a real constant (see (65.6)). The quantities c_1 and c_2 are real because they are coefficients in the solution ψ of a real equation (Schrödinger's equation) with real boundary conditions (the asymptotic form of the stationary wave as $r \to \infty$). The properties of the wave function ψ when inelastic processes are present differ only in that the condition imposed on ψ at infinity becomes complex: the asymptotic expression (72.1) with different amplitudes of the ingoing and outgoing waves does not reduce to a real stationary wave. The constant β therefore becomes complex also: $\beta = \beta' + i\beta''$. The modulus $|S_0|$ is no longer equal to unity; the condition $|S_0| < 1$ means that the imaginary part of β must be negative ($\beta'' < 0$).

Retaining only the first term in (72.3) and (72.4), and substituting (73.1), we find the cross-sections for elastic and inelastic scattering:

$$\sigma_e = 4\pi\,|\beta|^2, \tag{73.2}$$

$$\sigma_r = 4\pi\,|\beta''|/k. \tag{73.3}$$

Thus the elastic scattering cross-section is again independent of velocity, but the inelastic cross-section is inversely proportional to the particle velocity—the $1/v$ *law* (H. A. Bethe 1935). Consequently, as the velocity diminishes, inelastic processes become more and more important in comparison with elastic scattering.

The $1/v$ law can be derived in another less rigorous but more easily visualised manner. Let us suppose that the probability that a reaction will occur in a collision is proportional to the squared modulus of the wave function of the incident wave at $r = 0$. Physically, this statement expresses the fact that, for example, a slow neutron colliding with a nucleus can bring about a reaction only if it "penetrates" into the nucleus. The reaction cross-section is found by dividing $|\psi_{\text{inc}}(0)|^2$ by the incident current density (or, equivalently, by taking ψ_{inc} as normalised to unit current). For a plane wave normalised to unit current we have $|\psi_{\text{inc}}|^2 \sim 1/v$, which is the required result.

In this argument it is assumed that the value of $\psi_{\text{inc}}(0)$ can be calculated from the wave function (plane wave) unperturbed by the field. For this to be true, and hence for the $1/v$ law to be valid, it is necessary that the field $U(r)$ acting on the incident particle should decrease sufficiently rapidly with increasing distance.[†] In particular, the $1/v$ law is not valid for reactions between charged particles interacting according to Coulomb's law.

§74. Inelastic collisions between fast particles and atoms

When a fast particle collides with an atom, various inelastic processes (excitation or ionisation of the atom) can occur, as well as elastic scattering. These processes can be considered in the Born ap-

---

[†] It can be shown that U must decrease more rapidly than $1/r^2$.

proximation, as has been done in §70 for elastic scattering of fast electrons. Here it is assumed that the velocity of the fast particle is large in comparison with those of the atomic electrons.

As already mentioned in §70, when an electron collides with an atom, the centre-of-mass system of coordinates may be regarded as coincident with the laboratory system, in which the atom is at rest. Let $\mathbf{p}$ and $\mathbf{p}'$ again be the initial and final momenta of the electron, and m its mass. We also define the vector $\hbar\mathbf{q} = \mathbf{p}'-\mathbf{p}$, representing the momentum transfer from the electron to the atom. The quantity $\mathbf{q}$ plays an important role in the process, and largely determines the nature of the collision. We shall consider two limiting cases, those of collisions where the momentum transfer is either large or small in comparison with $\hbar/a$, where a represents the dimensions of the atom.

The inequality $qa \gg 1$ signifies that the momentum given to the atom is large compared with the original momentum of the atomic electrons. It is physically evident that in this case the atomic electrons may be regarded as free, and the collision of a fast electron with the atom may be regarded as an elastic collision with one of the atomic electrons, originally at rest. The cross-section for scattering by each of the Z electrons is given by Rutherford's formula; if the incident electron and the atomic electron acquire velocities of comparable magnitude, exchange effects become important, and the cross-section is determined by formula (69.7).

Let us now consider the opposite case of small momentum transfer ($qa \ll 1$). This means that the electron is deflected through a very small angle, and the energy transferred to the atom is small in comparison with its original energy. These properties enable us to take $p \approx p'$; then the vector $\mathbf{q}$ is simply the result of rotating $\mathbf{p}$ without changing its magnitude, and when the scattering angle ϑ is small we have

$$\hbar q \approx p\vartheta. \tag{74.1}$$

This expression becomes invalid only for very small angles: in the limit $\vartheta \to 0$, q tends to the value $q_{\min} = (p-p')/\hbar$, which depends on the small difference $p-p'$. The condition of energy conservation in the collision gives

$$E_n - E_0 = \frac{1}{2m}(p^2 - p'^2) \approx \frac{p}{m}(p - p') = v(p - p'),$$

where $E_n - E_0$ is the excitation energy of the atom when it goes from the ground level to the nth level, and v is the velocity of the incident electron. Hence the minimum value of the momentum transfer is

$$\hbar q_{\min} = (E_n - E_0)/v. \tag{74.2}$$

After this simplification, the only difference between the process under consideration and elastic scattering is that the initial and final states of the atom are not the same. The cross-section is therefore given by the same formula (70.1), except that ψ_0 and ψ_0^* in the integral become different wave functions ψ_0 and ψ_n^*:

$$d\sigma = \frac{m^2}{4\pi^2\hbar^4}\left|\iint Ue^{-i\mathbf{q}\cdot\mathbf{r}}\psi_n^*\psi_0 \, d\tau \, dV\right|^2. \tag{74.3}$$

The energy U includes the interaction of the incident electron with the nucleus of the atom and with all the Z atomic electrons:

$$U = -\frac{Ze^2}{r} + \sum_{a=1}^{Z} \frac{e^2}{|\mathbf{r} - \mathbf{r}_a|}, \tag{74.4}$$

where $\mathbf{r}$ and $\mathbf{r}_a$ are the radius vectors of the incident particle and the atomic electrons, and the origin is at the nucleus.

For inelastic processes, when (74.4) is substituted in (74.3), the term containing the interaction with the nucleus, Ze^2/r, disappears: the integration over τ in this term separates in the form $\int \psi_n^* \psi_0 \, d\tau$, which is zero because ψ_0 and ψ_n are orthogonal. In the other terms, the integration over V is carried out by means of the formula

$$\int \frac{e^{-i\mathbf{q}\cdot\mathbf{r}}}{|\mathbf{r} - \mathbf{r}_a|} dV = \frac{4\pi}{q^2} e^{-i\mathbf{q}\cdot\mathbf{r}_a} \tag{74.5}$$

(to derive which we need only note that the substitution $\mathbf{r} = \mathbf{r}_a + \mathbf{r}'$ reduces the integral to the form

$$e^{-i\mathbf{q}\cdot\mathbf{r}_a} \int e^{-i\mathbf{q}\cdot\mathbf{r}'} \, dV'/r' \equiv e^{-i\mathbf{q}\cdot\mathbf{r}_a}(1/r)_{\mathbf{q}},$$

and the Fourier component of $1/r$ is given by (68.6)). The result is

$$d\sigma_n = \left(\frac{2me^2}{\hbar^2}\right)^2 \left| \left(\sum_a e^{-i\mathbf{q}\cdot\mathbf{r}_a}\right)_{n0} \right|^2 \frac{do'}{q^4},$$

where the matrix element is taken with respect to the atomic wave functions:

$$\left(\sum_a e^{-i\mathbf{q}\cdot\mathbf{r}_a}\right)_{n0} = \sum_a \int e^{-i\mathbf{q}\cdot\mathbf{r}_a} \psi_n^* \psi_0 \, d\tau. \tag{74.6}$$

We can now make use of the fact that q is small. The variables $\mathbf{r}_a$ in the integral (74.6) have values lying in a volume with linear dimensions $\sim a$. Hence, when $qa \ll 1$, the quantities $\mathbf{q}.\mathbf{r}_a$ are small throughout this region, and we can put

$$e^{-i\mathbf{q}\cdot\mathbf{r}_a} \approx 1 - i\mathbf{q}.\mathbf{r}_a = 1 - iqx_a \tag{74.7}$$

(taking the direction of the vector $\mathbf{q}$ as the x-axis). Then

$$\left(\sum_a e^{-i\mathbf{q}\cdot\mathbf{r}_a}\right)_{n0} = -iq\left(\sum_a x_a\right)_{n0} = -i(q/e)\,(d_x)_{n0},$$

where $d_x = \sum ex_a$ is the Cartesian component of the dipole moment of the atom (the term unity disappears, since ψ_0 and ψ_n are orthogonal). Putting also

$$do' = 2\pi \sin\vartheta \, d\vartheta \approx 2\pi\vartheta \, d\vartheta = 2\pi(\hbar/mv)^2 q \, dq,$$

we get for the cross-section of the process

$$d\sigma_n = 8\pi(e/\hbar v)^2 \, |(d_x)_{n0}|^2 \, dq/q. \tag{74.8}$$

Thus the cross-section is determined by the squared matrix element of the dipole moment of the atom.[†]

[†] Here it is, of course, assumed that the matrix element does not vanish. If it does, the expansion (74.7) has to be continued to higher-order terms.

CHAPTER 11

PHOTONS

§75. The uncertainty principle in the relativistic case

The quantum theory described in Part I is essentially non-relativistic throughout, and is not applicable to phenomena involving motion at velocities comparable with that of light. At first sight, one might expect that the change to a relativistic theory is possible by a fairly direct generalisation of the formalism of non-relativistic quantum mechanics. But further consideration shows that this is not so.

We saw in Part I that quantum mechanics greatly restricts the possibility that an electron[†] simultaneously possesses values of different dynamical variables. For example, the uncertainties Δq and Δp in simultaneously existing values of the coordinate and the momentum are related by the expression $\Delta q \Delta p \sim \hbar$; the greater the accuracy with which one of these quantities is measured, the less the accuracy with which the other can be measured at the same time.

It is important to note, however, that any of the dynamical variables of the electron can individually be measured with arbitrarily high accuracy, and in an arbitrarily short period of time. This fact is of fundamental importance throughout non-relativistic quantum mechanics. It is the only justification for using the concept of the wave function $\psi(q)$, the square of whose modulus gives the probability of finding a particular value of the electron coordinate as the result of a measurement made at a given instant. The concept of such a prob-

---

[†] As in §1, we shall, for brevity, speak of an "electron", meaning any quantum system.

ability clearly requires that the coordinate can in principle be measured with any specified accuracy and rapidity, since otherwise this concept would be purposeless and devoid of physical significance.

The existence of a limiting velocity (the velocity of light, denoted by c) leads to new fundamental limitations on the possible measurements of various physical quantities (L. D. Landau and R. E. Peierls 1930).

In §37 the following relationship has been derived:

$$(v'-v)\,\Delta p\,\Delta t \sim \hbar, \tag{75.1}$$

relating the uncertainty Δp in the measurement of the electron momentum and the duration Δt of the measurement process itself; v and v' are the velocities of the electron before and after the measurement. From this relationship it follows that a momentum measurement of high accuracy made during a short time (i.e. with Δp and Δt both small) can occur only if there is a large change in the velocity as a result of the measurement process itself. In the non-relativistic theory, this showed that the measurement of momentum cannot be repeated at short intervals of time, but it did not all diminish the possibility, in principle, of making a single measurement of the momentum with arbitrarily high accuracy, since the difference $v'-v$ could take any value, no matter how large.

The existence of a limiting velocity, however, radically alters the situation. The difference $v'-v$, like the velocities themselves, cannot now exceed c (or rather $2c$). Replacing $v'-v$ in (75.1) by c, we obtain

$$\Delta p\,\Delta t \sim \hbar/c, \tag{75.2}$$

which determines the highest accuracy theoretically attainable when the momentum is measured by a process occupying a given time Δt. In the relativistic theory, therefore, it is in principle impossible to make an arbitrarily accurate and rapid measurement of the momentum. An exact measurement ($\Delta p \to 0$) is possible only in the limit as the duration of the measurement tends to infinity.

Equally fundamental changes occur in regard to the measurability of the coordinate. In the relativistic theory, the coordinate cannot be measured with an accuracy better than a certain limit. The concept of localising the electron is thereby further restricted in its physical significance.

In the mathematical formalism of the theory, this situation is shown by the fact that an accurate measurement of the coordinate is incompatible with the assertion that the energy of a free particle is positive. It will be seen later that the complete set of eigenfunctions of the relativistic wave equation of a free particle includes, as well as solutions having the "correct" time dependence, also solutions having a "negative frequency". These functions will in general appear in the expansion of the wave packet corresponding to an electron localised in a small region of space.

It will be shown that the wave functions having a "negative frequency" correspond to the existence of antiparticles (positrons). The appearance of these functions in the expansion of the wave packet expresses the (in general) inevitable production of electron–positron pairs in the process of measuring the coordinates of an electron. This formation of new particles in a way which cannot be detected by the process itself clearly renders meaningless the measurement of the electron coordinates.

In the rest frame of the electron, the least possible error in the measurement of its coordinates is

$$\Delta q \sim \hbar/mc. \tag{75.3}$$

This value (which purely dimensional arguments show to be the only possible one) corresponds to a momentum uncertainty $\Delta p \sim mc$, which in turn corresponds to the threshold energy for pair production.

In a frame of reference in which the electron is moving with energy ε, (75.3) becomes

$$\Delta q \sim \hbar c/\varepsilon. \tag{75.4}$$

In particular, in the limiting ultra-relativistic case the energy is related to the momentum by $\varepsilon = cp$, and

$$\Delta q \sim \hbar/p, \tag{75.5}$$

i.e. the error Δq is the same as the de Broglie wavelength of the particle.

It is clear from the foregoing that, in a consistent relativistic quantum mechanics, the coordinates of particles cannot act as dynamical variables, since these must by their nature have a precise significance. Nor can the particle momentum retain its former meaning. Since an accurate measurement of the momentum requires a long interval of time, there is no possibility of following its change in the process.

Having regard to the discussion at the beginning of this section, we reach the conclusion that the entire formalism of non-relativistic quantum mechanics becomes insufficient in the relativistic case. The wave functions $\psi(q)$, in their original sense as the carriers of unobservable information, cannot appear in the formalism of a consistent relativistic theory.

The momentum can figure in a consistent theory only for free particles; for these it is conserved, and can therefore be measured with any desired accuracy. This indicates that the theory will not consider the time dependence of particle interaction processes. The only observable quantities are the properties (momenta, polarisations) of free particles: the initial particles which come into interaction, and the final particles which result from the process.

A typical problem as formulated in relativistic quantum theory is to determine the probability amplitudes of transitions between specified initial and final states of a system of particles. The set of such amplitudes between all possible states constitutes the *scattering matrix* or *S-matrix*. This matrix will embody all the information about particle interaction processes that has an observable physical meaning (W. Heisenberg 1938).

In such a theory, moreover, the concepts of "elementary" and "composite" particles lose their earlier significance; the problem of "what consists of what" cannot be formulated without considering the process of interaction between particles, and if this is not done the whole problem becomes meaningless. All particles which occur as initial or final particles in any physical collision phenomenon must appear in the theory on an equal footing. In this sense the difference between those particles usually said to be "composite" and those said to be "elementary" is only a quantitative one, and amounts to the value of the mass defect with respect to decay into specified "component parts". For example, the statement that the deuteron is composite (its binding energy with respect to disintegration into a proton and a neutron being fairly small) differs only quantitatively from the statement that the neutron "consists of" a proton and a pion.

There is as yet no logically consistent and complete relativistic quantum theory. We shall see that the existing theory introduces new physical features into the nature of the description of particle states, which acquires some of the features of field theory. The theory is, however, largely constructed on the pattern of ordinary quantum mechanics and makes use of the latter's concepts. This structure of the theory has yielded good results in quantum electrodynamics. The lack of complete logical consistency in this theory is shown by the occurrence of divergent expressions when the mathematical formalism is directly applied, although there are quite well-defined ways of eliminating these divergences. Nevertheless, such methods remain, to a considerable extent, semiempirical rules, and our confidence in the correctness of the results is ultimately based only on their excellent agreement with experiment, not on the internal consistency or logical ordering of the fundamental principles of the theory.

A quite different situation occurs in the theory of effects depending on the strong interactions of particles (nuclear forces). Here, attempts to construct a theory by similar methods have led to no systematic results bearing on physical reality. The construction of a complete theory embracing strong interactions will probably call for the application of fundamentally new physical ideas.

§76. Quantisation of the free electromagnetic field[†]

A natural means of transition from the classical to the quantum description of the electromagnetic field consists in the classical expansion of the field in terms of oscillators. The basis of this procedure is as follows (cf. *Mechanics and Electrodynamics*, §78).

We describe a free electromagnetic field (or electromagnetic waves) by potentials taken in a gauge such that the scalar potential is zero, leaving only the vector potential **A**. Considering the field in some large but finite volume Ω, we can expand it in travelling plane waves, and the potential is then represented by a series

$$\mathbf{A} = \sum_{\mathbf{k}} \sqrt{\frac{2\pi}{\omega\Omega}} (\mathbf{c}_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} + \mathbf{c}_{\mathbf{k}}^{*} e^{-i\mathbf{k}\cdot\mathbf{r}}), \tag{76.1}$$

where the coefficient $\mathbf{c}_{\mathbf{k}}$ are functions of the time such that

$$\mathbf{c}_{\mathbf{k}} \sim e^{-i\omega t}, \quad \omega = |\mathbf{k}|, \tag{76.2}$$

and each is orthogonal to the corresponding wave vector: $\mathbf{c}_{\mathbf{k}} . \mathbf{k} = 0$.[‡] The summation in (76.1) is taken over an infinite but discrete set of adjacent values of the vector (i.e. of its three components k_x, k_y, k_z). The change from summation to integration over a continuous distribution may be made by means of the expression

$$\Omega \, dk_x \, dk_y \, dk_z/(2\pi)^3 \tag{76.3}$$

for the number of possible values of **k** belonging to the volume element in **k**-space.

[†] From here onwards in Chapters 11–16, except where otherwise specified, we shall use *relativistic units*, in which the velocity of light c and Planck's constant $\hbar$ are taken as unity; this considerably simplifies the formulae. In these units, energy, momentum, and mass have the same dimensions—those of reciprocal length. The square of the unit charge is equal to the value of the dimensionless (in ordinary units) constant $e^2/\hbar c$, or 1/137.

[‡] The definition of the coefficients $\mathbf{c}_{\mathbf{k}}$ in (76.1) differs from that of the $\mathbf{a}_{\mathbf{k}}$ in *Mechanics and Electrodynamics* (78.1) by a factor $\sqrt{(\omega\Omega/2\pi)}$. The convenience of this definition in making the change to the quantum theory will be explained below.

If the vectors $\mathbf{c}_\mathbf{k}$ are specified, the field in the volume considered is completely determined. Thus these quantities may be regarded as a discrete set of classical "field variables". In order to explain the transition to the quantum theory, however, a further transformation of these variables is needed, whereby the field equations take a form analogous to the canonical equations (Hamilton's equations) of classical mechanics. The canonical field variables are defined as the real quantities

$$\left.\begin{aligned} \mathbf{Q}_\mathbf{k} &= \frac{1}{\sqrt{(2\omega)}}(\mathbf{c}_\mathbf{k}+\mathbf{c}_\mathbf{k}^*),\\ \mathbf{P}_\mathbf{k} &= -\frac{i\omega}{\sqrt{(2\omega)}}(\mathbf{c}_\mathbf{k}-\mathbf{c}_\mathbf{k}^*) = \dot{\mathbf{Q}}_\mathbf{k}. \end{aligned}\right\} \tag{76.4}$$

The Hamiltonian (i.e. the energy) of the field is given in terms of these variables by

$$H = \tfrac{1}{2}\sum_\mathbf{k}(\mathbf{P}_\mathbf{k}^2+\omega^2\mathbf{Q}_\mathbf{k}^2).$$

Each of the vectors $\mathbf{P}_\mathbf{k}$ and $\mathbf{Q}_\mathbf{k}$ is perpendicular to the wave vector $\mathbf{k}$, and therefore has two independent components. The direction of these vectors determines the direction of polarisation of the corresponding wave. Denoting the components of the vectors $\mathbf{P}_\mathbf{k}$, $\mathbf{Q}_\mathbf{k}$ (in the plane perpendicular to $\mathbf{k}$) by $P_{\mathbf{k}\sigma}$, $Q_{\mathbf{k}\sigma}$ ($\sigma = 1, 2$), we can write the Hamiltonian as

$$H = \sum_{\mathbf{k},\sigma} H_{\mathbf{k}\sigma}, \quad H_{\mathbf{k}\sigma} = \tfrac{1}{2}(P_{\mathbf{k}\sigma}^2+\omega^2 Q_{\mathbf{k}\sigma}^2). \tag{76.5}$$

Thus the Hamiltonian is the sum of independent terms, each of which contains only one pair of the quantities $P_{\mathbf{k}\sigma}$, $Q_{\mathbf{k}\sigma}$. Each such term corresponds to a travelling wave with a definite wave vector and polarisation, and has the form of the Hamiltonian for a one-dimensional harmonic oscillator.

This classical description of the field makes the manner of transition to the quantum theory obvious. We have now to use the canonical variables (generalised coordinates $Q_{\mathbf{k}\sigma}$ and generalised momenta $P_{\mathbf{k}\sigma}$) as operators, with the usual commutation rules for coordinates and momenta:

$$\hat{P}_{\mathbf{k}\sigma}\hat{Q}_{\mathbf{k}\sigma}-\hat{Q}_{\mathbf{k}\sigma}\hat{P}_{\mathbf{k}\sigma} = -i; \tag{76.6}$$

operators with different values of **k** and σ always commute. The field potential **A** is then likewise an operator.

The Hamiltonian of the field is found by replacing the canonical variables in (76.5) by the corresponding operators:

$$\hat{H} = \sum_{\mathbf{k},\sigma} \hat{H}_{\mathbf{k}\sigma}, \qquad \hat{H}_{\mathbf{k}\sigma} = \tfrac{1}{2}(\hat{P}^2_{\mathbf{k}\sigma}+\omega^2\hat{Q}^2_{\mathbf{k}\sigma}). \tag{76.7}$$

The determination of the eigenvalues of this Hamiltonian involves no further calculation, since it is equivalent to the problem of the energy levels of linear oscillators, the solution of which we already know (§25). We can therefore immediately write down the field energy levels:

$$E = \sum_{\mathbf{k},\sigma} (N_{\mathbf{k}\sigma}+\tfrac{1}{2})\omega, \tag{76.8}$$

where the $N_{\mathbf{k}\sigma}$ are integers.

The classical expression for the field momentum is

$$\mathbf{P} = \sum_{\mathbf{k},\sigma} \mathbf{n}H_{\mathbf{k}\sigma},$$

where $\mathbf{n} = \mathbf{k}/k$; see *Mechanics and Electrodynamics* (78.12). The corresponding operator is found by replacing $H_{\mathbf{k}\sigma}$ by $\hat{H}_{\mathbf{k}\sigma}$, and its eigenvalues are therefore

$$\mathbf{P} = \sum_{\mathbf{k},\sigma} \mathbf{k}(N_{\mathbf{k}\sigma}+\tfrac{1}{2}). \tag{76.9}$$

The further discussion of formulae (76.8) and (76.9) will be left until §77; here we shall write out the matrix elements of the quantities $Q_{\mathbf{k}\sigma}$, which can be done at once by means of the known formulae (25.4) for the matrix elements of the coordinate of an oscillator. The non-zero matrix elements are

$$\langle N_{\mathbf{k}\sigma}\,|Q_{\mathbf{k}\sigma}|\,N_{\mathbf{k}\sigma}-1\rangle = \langle N_{\mathbf{k}\sigma}-1\,|Q_{\mathbf{k}\sigma}|\,N_{\mathbf{k}\sigma}\rangle = \sqrt{(N_{\mathbf{k}\sigma}/2\omega)}. \tag{76.10}$$

The matrix elements of the quantities $P_{\mathbf{k}\sigma} = \dot{Q}_{\mathbf{k}\sigma}$ differ from those of $Q_{\mathbf{k}\sigma}$ only by the factor $\pm i\omega$, according to the general rule (11.8):

$$\langle N_{\mathbf{k}\sigma}\,|P_{\mathbf{k}\sigma}|\,N_{\mathbf{k}\sigma}-1\rangle = -\langle N_{\mathbf{k}\sigma}-1\,|P_{\mathbf{k}\sigma}|\,N_{\mathbf{k}\sigma}\rangle = i\omega\sqrt{(N_{\mathbf{k}\sigma}/2\omega)}.$$

As will be seen later, however, a more fundamental significance attaches not to the operators $\hat{Q}_{\mathbf{k}\sigma}$ and $\hat{P}_{\mathbf{k}\sigma}$ themselves but to linear combinations of them, given by

$$\left.\begin{aligned} \hat{c}_{\mathbf{k}\sigma} &= \frac{1}{\sqrt{(2\omega)}}(\omega\hat{Q}_{\mathbf{k}\sigma}+i\hat{P}_{\mathbf{k}\sigma}), \\ \hat{c}^{+}_{\mathbf{k}\sigma} &= \frac{1}{\sqrt{(2\omega)}}(\omega\hat{Q}_{\mathbf{k}\sigma}-i\hat{P}_{\mathbf{k}\sigma}), \end{aligned}\right\} \tag{76.11}$$

which correspond to the definition of the coefficients $\hat{c}_{\mathbf{k}\sigma}$ in the classical expansion (76.1). The only non-zero matrix elements of these operators are

$$\langle N_{\mathbf{k}\sigma}-1\,|c_{\mathbf{k}\sigma}|\,N_{\mathbf{k}\sigma}\rangle = \langle N_{\mathbf{k}\sigma}\,|c^{+}_{\mathbf{k}\sigma}|\,N_{\mathbf{k}\sigma}-1\rangle = \sqrt{N_{\mathbf{k}\sigma}}. \tag{76.12}$$

Using the definition (76.11) and the rule (76.6), we can easily find the commutation rule for the operators $\hat{c}_{\mathbf{k}\sigma}$ and $\hat{c}^{+}_{\mathbf{k}\sigma}$:

$$\hat{c}_{\mathbf{k}\sigma}\hat{c}^{+}_{\mathbf{k}\sigma}-\hat{c}^{+}_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}\sigma} = 1. \tag{76.13}$$

Thus we obtain the electromagnetic field operator in the form

$$\hat{\mathbf{A}} = \sum_{\mathbf{k},\sigma}\sqrt{\frac{2\pi}{\omega\Omega}}\left(\hat{c}_{\mathbf{k}\sigma}\mathbf{e}^{(\sigma)}e^{i\mathbf{k}.\mathbf{r}}+\hat{c}^{+}_{\mathbf{k}\sigma}\mathbf{e}^{(\sigma)*}e^{-i\mathbf{k}.\mathbf{r}}\right). \tag{76.14}$$

Here $\mathbf{e}^{(\sigma)}$ denotes the unit vectors in the direction of polarisation of the oscillators; these vectors are perpendicular to the wave vectors **k**, and for every k there are two independent polarisations denoted by the superscript $\sigma = 1, 2$.[†]

The expression (76.14) corresponds to the usual representation of operators in non-relativistic quantum theory, which has been tacitly used throughout. In this *Schrödinger representation*, the operators of the various physical quantities do not themselves depend explicitly on

---

[†] For linear polarisation, the unit vector **e** is real, and gives the direction of polarisation immediately (see *Mechanics and Electrodynamics*, §70). For circular (or, more generally, elliptical) polarisation the vector **e** is complex, with a definite ratio of the real and imaginary parts; here the description "unit vector" must be taken as meaning that $\mathbf{e}.\mathbf{e}^* = 1$.

the time. The variation of the system with time is described by the time dependence of the wave function. The formalism of quantum mechanics can also, however, be expressed in a slightly different but equivalent form, in which the explicit time dependence is transferred from the wave functions to the operators; this is called the *Heisenberg representation* of the operators. Such a formulation is especially convenient for the description of fields in relativistic quantum theory: the dependence of the operators on coordinates and time placed on an equal footing allows a clearer elucidation of the relativistic space-time invariance of the theory, whereas in the Schrödinger formulation the space coordinates and the time appear in a highly asymmetric manner.

For the operator $\hat{\mathbf{A}}$, the change to the Heisenberg representation amounts to including a factor $e^{-i\omega t}$ (or its conjugate) in each term of the sum (76.14), corresponding to the time dependence of the "stationary states of the field oscillators". The final expression for the operator $\hat{\mathbf{A}}$ may be written

$$\hat{\mathbf{A}}(r, t) = \sum_{\mathbf{p},\sigma} (\hat{c}_{\mathbf{k}\sigma}\mathbf{A}_{\mathbf{k}\sigma} + \hat{c}^{+}_{\mathbf{k}\sigma}\mathbf{A}^{*}_{\mathbf{k}\sigma}), \tag{76.15}$$

where

$$\mathbf{A}_{\mathbf{k}\sigma} = \mathbf{e}^{(\sigma)}\sqrt{\frac{2\pi}{\omega\Omega}}\, e^{-i(\omega t - \mathbf{k}\cdot\mathbf{r})}. \tag{76.16}$$

Henceforward, in the discussion of either electromagnetic fields or particle fields, we shall use the Heisenberg representation of the field operators.

§77. **Photons**

We shall now further analyse the field quantisation formulae obtained in §76.

First of all, formula (76.8) for the field energy raises the following difficulty. The lowest energy level of the field corresponds to the case where the quantum numbers $N_{\mathbf{k}\sigma}$ of all the oscillators are zero; this is called the *electromagnetic field vacuum* state. But, even in that state, each oscillator has a non-zero "zero-point energy" equal to $\frac{1}{2}\omega$. Summation over an infinite number of oscillators then gives an infinite result. Thus we meet with one of the "divergences" which are due to the fact that the present theory is not logically complete and consistent.

So long as only the field energy eigenvalues are under discussion, we can remove this difficulty by simply striking out the zero-point oscillation energy, i.e. by writing the field energy and momentum (in ordinary units) as

$$E = \sum_{\mathbf{k},\sigma} N_{\mathbf{k}\sigma}\hbar\omega, \qquad \mathbf{P} = \sum_{\mathbf{k},\sigma} N_{\mathbf{k}\sigma}\hbar\mathbf{k}. \tag{77.1}$$

These formulae enable us to introduce the concept of *radiation quanta* or *photons*, which is fundamental throughout quantum electrodynamics.[†] We may regard the free electromagnetic field as an ensemble of particles each with energy $\hbar\omega$ and momentum $\hbar\mathbf{k} = \mathbf{n}\hbar\omega/c$. The relationship between the energy and momentum is as it should be in relativistic mechanics for particles having zero rest-mass and moving with the velocity of light. The numbers $N_{\mathbf{k}\sigma}$ now represent the numbers of photons having given momentum $\mathbf{k}$ and polarisation $\mathbf{e}^{(\sigma)}$. The polarisation of the photon is analogous to the spin of other particles; the exact properties of the photon in this respect will be discussed in §78 below.

It is easily seen that the whole of the mathematical formalism developed in §76 is fully in accordance with the representation of the free electromagnetic field as an ensemble of photons; it is just the second quantisation formalism, applied to the system of photons. In this treatment (see §47), the independent variables are the occupation numbers of the states (in the present case, the numbers $N_{\mathbf{k}\sigma}$), and the operators act on functions of these numbers. The particle "annihilation" and "creation" operators are of basic importance; they respectively decrease and increase by one the occupation numbers. The $\hat{c}_{\mathbf{k}\sigma}$ and $\hat{c}^{+}_{\mathbf{k}\sigma}$ are operators of this kind: $\hat{c}_{\mathbf{k}\sigma}$ annihilates a photon in the state $\mathbf{k}$, σ (and has matrix elements only for the transitions $N_{\mathbf{k}\sigma} \to N_{\mathbf{k}\sigma}-1$; cf. (76.12)); $\hat{c}^{+}_{\mathbf{k}\sigma}$ creates a photon in that state (its matrix elements being non-zero only for transitions $N_{\mathbf{k}\sigma} \to N_{\mathbf{k}\sigma}+1$).

[†] This concept is originally due to A. Einstein (1905), in connection with the theory of the photoelectric effect.

The plane waves (76.16) which appear in the operator (76.15) as coefficients of the photon annihilation operators may be treated as the wave functions of photons having given momenta **k** and polarisations $\mathbf{e}^{(\sigma)}$, normalised to one photon in the volume Ω. This corresponds to an expansion (47.22) of the ψ-operator in terms of the wave functions of stationary states of a particle in the non-relativistic second quantisation formalism; however, unlike the latter, the expansion (76.15) includes both particle annihilation and particle creation operators. The meaning of this difference is explained in Chapter 13.

Here it should be emphasised once more that the "wave function" of the photon cannot be regarded as the probability amplitude of its spatial localisation, in contrast to the fundamental significance of the wave function in non-relativistic quantum mechanics. This is especially clear in the case of the photon, which is always ultra-relativistic, so that the minimum error in its coordinates is $\Delta q \sim 1/k \sim \lambda$, by (75.5). This means that one can speak of the coordinates of the photon only when the characteristic dimensions of the problem are large compared with the wavelength. This is just the "classical" limiting case corresponding to geometrical optics, in which one can regard radiation as being propagated along definite paths or rays. In the quantum case, however, when the wavelength cannot be regarded as small, the concept of the coordinates of the photon becomes meaningless.

The commutation rule (76.13) for the photon creation and annihilation operators corresponds to the case of particles that obey Bose statistics; cf. (47.11). Photons are therefore bosons. In accordance with the properties of this statistics, there can be any number of photons simultaneously present in any given state.

The description of the field as an ensemble of photons is the only one that is fully adequate to the physical significance of the free electromagnetic field in quantum theory. It replaces the classical description based on field potentials (and field strengths). These appear as second quantisation operators in the mathematical treatment of the photon picture.

The properties of a quantum system become almost classical when the quantum numbers defining the stationary states of the system are large (§27). For a free electromagnetic field (in a given volume), this means that the quantum numbers of the oscillators, i.e. the photon numbers $N_{\mathbf{k}\sigma}$, must be large. In this sense the fact that photons obey Bose statistics is of great importance. In the mathematical formalism of the theory, the connection of Bose statistics with the properties of the classical field is shown by the commutation rules for the operators $\hat{c}_{\mathbf{k}\sigma}$, $\hat{c}^{+}_{\mathbf{k}\sigma}$. When the $N_{\mathbf{k}\sigma}$ are large, and the matrix elements of these operators are also large, the right-hand side of the commutation relation (76.13) can be neglected, and we then have $\hat{c}_{\mathbf{k}\sigma}\hat{c}^{+}_{\mathbf{k}\sigma} = \hat{c}^{+}_{\mathbf{k}\sigma}\hat{c}_{\mathbf{k}\sigma}$, i.e. these operators become the commuting classical quantities $c_{\mathbf{k}\sigma}$, $c^{*}_{\mathbf{k}\sigma}$ which define the classical field potentials.

§78. The angular momentum and parity of the photon

The photon, like any other particle, can possess a certain angular momentum, but the properties of this angular momentum differ somewhat from those for ordinary particles. To see the reason for this difference, let us first recall the relationship between the properties of the wave function of a particle and the angular momentum of the particle, in the mathematical formalism of quantum mechanics.

The angular momentum $\mathbf{j}$ of a particle consists of its orbital angular momentum $\mathbf{l}$ and its intrinsic angular momentum or spin $\mathbf{s}$. The wave function of a particle having spin s is a symmetrical spinor of rank $2s$, i.e. is a set of $2s+1$ components which are transformed into definite combinations of one another when the coordinate axes are rotated (§41). The orbital angular momentum is related to the way in which the wave functions depend on the coordinates: states with orbital angular momentum l correspond to wave functions whose components are linear combinations of the spherical harmonic functions of order l.

The vector $\mathbf{A}$ plays the part of the wave function of the photon. A vector is equivalent to a spinor of rank two, and in this sense the photon may be assigned spin 1. Since this is integral, it follows in turn that the total angular momentum of the photon also can take only integral values: $j = 1, 2, 3, \ldots$. There is no value $j = 0$ for the photon: the wave function of a state with zero angular momentum must be spherically symmetrical, and this certainly cannot occur for a transverse wave.

Whereas the total angular momentum of the photon has an entirely precise significance, the photon spin in only conventional: one cannot consistently distinguish the spin and the orbital angular momentum of the photon as constituent parts of its total angular momentum. The reason is that it would then be necessary for the "spin" and "coordinate" properties of the wave functions to be independent: the coordinate dependence of the components of a spinor (in this case a vector) cannot be limited by any extra conditions; but the vector wave function **A** of the photon is subject to the further condition of transversality, so that the coordinate dependence cannot be simultaneously and arbitrarily specified for all its components. The definition of the spin as the angular momentum of a particle at rest is also inapplicable to the photon, since there is no rest frame for the photon, which moves at the speed of light.

The state of the photon, like that of any particle, can also be described by its parity, which refers to the behaviour of the wave function under inversion of the coordinates. The state is said to be *even* if the vector wave function $\mathbf{A}(\mathbf{r})$ is unchanged by inversion, and *odd* if $\mathbf{A}(\mathbf{r})$ changes sign.[†] There is an accepted terminology for the various states of a photon that have definite angular momenta and parities: a photon in a state with angular momentum j and parity $(-1)^j$ is called an *electric* 2^j*-pole* (or *Ej*) *photon*; one with parity $(-1)^{j+1}$ is called a *magnetic* 2^j*-pole* (or *Mj*) *photon*.[‡] The angular momentum and parity of a particle are frequently denoted by a single symbol in which a

---

[†] The effect of inversion on a scalar function $\phi(\mathbf{r})$ is to change the sign of the argument: $\hat{P}\phi(\mathbf{r}) = \phi(-\mathbf{r})$. When the inversion operator acts on a vector function $\mathbf{A}(\mathbf{r})$, it must also be borne in mind that the reversal of the direction of the coordinate axes also changes the sign of each component of the (polar) vector. Thus the effect of the inversion operation is shown by $\hat{P}\mathbf{A}(\mathbf{r}) = -\mathbf{A}(-\mathbf{r})$. Hence, for example, in an even state we must have $\mathbf{A}(-\mathbf{r}) = -\mathbf{A}(\mathbf{r})$ if $\hat{P}\mathbf{A}(\mathbf{r}) = \mathbf{A}(\mathbf{r})$.

[‡] These names are in accordance with the terminology of radiation theory: the emission of electric and magnetic photons is respectively caused by the electric and magnetic multipole moments of a system of charges (see §98).

number shows the value of j and a superscript + or − denotes the parity $P = +1$ or -1. Thus electric photons correspond to states 1^-, 2^+, 3^-, 4^+, etc., and magnetic photons to 1^+, 2^-, 3^+, 4^-, etc. In particular, an electric dipole photon corresponds to a 1^- state and a magnetic dipole photon to a 1^+ state.

The state of a photon with a definite value of j is a spherical wave in which there is no distinctive direction of motion. On the other hand, if a photon has a definite direction of motion (i.e. has a definite momentum vector $\mathbf{k}$), then it has no definite value of j. A photon with a definite direction of $\mathbf{k}$ can, however, also have a definite value of the angular momentum component in that direction; this component is called the *helicity* and denoted by λ.[†]

The conservation of helicity, like that of any component of the angular momentum, is due to certain symmetry properties of space in relation to a free particle. The momentum $\mathbf{k}$ distinguishes a particular direction in space. The existence of this direction removes the complete symmetry with respect to arbitrary rotations of the coordinate axes (and so the angular momentum vector is no longer conserved). There is still, however, axial symmetry with respect to rotations about the selected axis (the direction of $\mathbf{k}$). The conservation of helicity expresses this symmetry.

By the definition of the orbital angular momentum operator $\hat{\mathbf{l}} = \mathbf{r} \times \hat{\mathbf{p}}$, the operator of the component of this angular momentum in the direction of the momentum is identically zero, as are the eigenvalues of this component. The helicity is therefore equal to the component of the particle spin in its direction of motion. For an ordinary particle with spin 1, the helicity may therefore have the values 0 and ± 1. For the photon, as we shall now show, only the values $\lambda = \pm 1$ are possible; this again shows that the concept of photon spin is a purely conventional one.

It is easily seen that the states of the photon having definite helicities are the same as its states of circular polarisation. Let ξ, η, ζ be coordi-

---

[†] This is to be distinguished from m, the component of the angular momentum in a fixed direction in space (the z-axis).

nates with the ζ-axis in the direction of the photon momentum (whereas the position of the z-axis is unaffected by the motion of the particle). Let us consider, for instance, a photon state with helicity $\lambda = +1$. According to formulae (41.9), which give the relation between the components of a vector wave function (of a particle with spin 1) and the components of a spinor of rank two, such a state corresponds to a wave function **A** whose components are related by $A_\eta = iA_\xi$, $A_\zeta = 0$; for in that case the only non-zero component of the spinor is ψ^{11}, which corresponds to the value $+1$ for the ζ-component of the spin. Similarly, a wave function with components $A_\eta = -iA_\xi$, $A_\zeta = 0$ corresponds to $\lambda = -1$. Such relationships are satisfied not only by the vector **A** but also by the polarisation vector **e**, which appears as a factor in (76.16). The values $e_\eta = \pm ie_\xi$, however, correspond to circular polarisation (see *Mechanics and Electrodynamics*, §70).

The impossibility of the value $\lambda = 0$ is evident from the fact that this would have to correspond to a wave function with components $A_\xi = A_\eta = 0$, $A_\zeta \neq 0$, which, according to (41.9), is equivalent to the spinor component ψ^{12}; but this is excluded by the requirement that the vector **A** is transverse to **k**.

CHAPTER 12

DIRAC'S EQUATION

§79. The Klein–Fock equation

We shall begin the discussion of the relativistic quantum theory of particles by considering the properties of wave functions describing particles and by constructing the wave equation that is satisfied by these functions. In the non-relativistic theory, the wave functions of particles with different spins are spinors of different ranks, and the wave functions of free particles all satisfy the same equation, namely Schrödinger's equation for free motion. In the relativistic theory, however, as we shall see, the form of the wave equation of free motion depends essentially on the particle spin.

The simplest case is, of course, that of particles with spin zero. In the non-relativistic theory, they are described by scalar wave functions. In the relativistic theory, the three-dimensional scalar becomes a four-dimensional scalar that is invariant not only under transformations of the spatial coordinates but also under Lorentz transformations.

In relativistic mechanics, the energy ε of a particle and its momentum $\mathbf{p}$ form a four-vector $p^{\mu} = (\varepsilon, \mathbf{p})$.[†] Accordingly, the operators corresponding to these quantities also form a four-vector $\hat{p}^{\mu}$. The operator $\hat{\mathbf{p}} = -i\nabla$ corresponds to the three-dimensional momentum $\mathbf{p}$, and the operator $i\,\partial/\partial t$ of differentiation with respect to time corresponds to the energy (the Hamiltonian) in the wave equation; cf. (8.1).

[†] In Chapters 12–16 we shall denote by ε the relativistic energy of an individual particle, including the rest energy.

Thus the four-momentum operator is

$$\hat{p}^\mu = (i\,\partial/\partial t, -i\nabla), \quad \hat{p}_\mu = (i\,\partial/\partial t, i\nabla) \tag{79.1}$$

or, in four-dimensional notation,

$$\hat{p}_\mu = i\,\partial/\partial x^\mu. \tag{79.2}$$

If the scalar operator $\hat{p}_\mu \hat{p}^\mu$, the square of the four-vector $\hat{p}^\mu$, acts on any wave function Ψ, the square of the four-momentum is just a constant, the square of the mass m of the particle, and so the result is to multiply the wave function by m^2:

$$\hat{p}_\mu \hat{p}^\mu \Psi = m^2 \Psi, \tag{79.3}$$

which can be expanded as

$$\left(-\frac{\partial^2}{\partial t^2} + \triangle\right)\Psi = m^2\Psi \tag{79.4}$$

(O. Klein and V. A. Fock 1926).

For a relativistic particle with spin zero, there is no Hamiltonian in the sense defined in non-relativistic theory: equation (79.4) is of the second order with respect to time, whereas the Hamiltonian $\hat{H}$ would have to determine the first derivative of the wave function according to $i\,\partial\Psi/\partial t = \hat{H}\Psi$.

For a spin-zero particle, the probability density of its various positions in space certainly could not be determined by the squared modulus $|\Psi|^2$, even formally (quite apart from the general physical considerations given in §75, which prevent the wave function from being regarded as carrying information about the spatial localisation of the particle). The reason is that in the relativistic theory the particle distribution and current densities form a four-vector (cf. the discussion of the current density vector in *Mechanics and Electrodynamics*, §54). The particle density is the time component of this four-vector, and not a scalar. It therefore cannot be determined by a scalar quantity such as the squared modulus of a scalar function.

For reasons to be explained later (§92), the description of particles by means of the scalar wave equation (79.4) has only a very restricted significance. We shall therefore not pause here to elucidate the mathematical structure of the quantities which play the part of the particle current density and energy density four-vector.

§80. **Four-dimensional spinors**

In non-relativistic quantum theory, a particle with spin s is described by a symmetrical spinor of rank $2s$, i.e. a set of $2s+1$ quantities that are transformed into particular combinations of one another when the coordinate axes are rotated. The law of this transformation represents the symmetry properties of the particles resulting from the isotropy of space.

In the relativistic theory, rotations of the space coordinates occur only as a special case of four-dimensional rotations (rotations of the four-dimensional space–time coordinates). The set of all possible transformations of this kind is called the *Lorentz group*. It includes not only the three-dimensional rotations, which leave the direction of the time axis unaltered, but also the ordinary Lorentz transformations, which are rotations in the *xt*, *yt* or *zt* plane (see *Mechanics and Electrodynamics*, §36). A general four-dimensional rotation is a Lorentz transformation together with a rotation of the space coordinates.

To describe particles with spin in relativistic quantum theory, it is therefore necessary to develop the theory of four-dimensional spinors (four-spinors), which play the same part with respect to Lorentz-group transformations as the ordinary (three-dimensional) spinors do with respect to the space-rotation group.[†]

A four-spinor of rank one:

$$\xi = \begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix} \tag{80.1}$$

is a quantity with two components, which is transformed by any

---

[†] That is, the four-spinors provide irreducible representations of the Lorentz group, just as the three-dimensional spinors provide irreducible representations of the rotation group.

Lorentz-group transformation in accordance with formulae similar to (41.3):

$$\xi^{1\prime} = \alpha\xi^1+\beta\xi^2, \quad \xi^{2\prime} = \gamma\xi^1+\delta\xi^2; \tag{80.2}$$

the complex coefficients $\alpha, \beta, \gamma, \delta$ are now definite functions of the angles of rotation of the four-dimensional coordinate axes (in general, there are six such angles, corresponding to rotations in the six coordinate planes xy, xz, yz, tx, ty, tz). Being components of the wave function of a particle with spin $\frac{1}{2}$, ξ^1 and ξ^2 correspond to the eigenvalues of the z-projection of the spin, which are $+\frac{1}{2}$ and $-\frac{1}{2}$.

For the same reason as with three-dimensional spinors, the coefficients in the transformation (80.2) are related by (41.5):

$$\alpha\delta-\gamma\beta = 1. \tag{80.3}$$

This ensures the invariance of the bilinear antisymmetric combination

$$\xi^1\Xi^2-\xi^2\Xi^1 \tag{80.4}$$

of the components of any two spinors ξ and Ξ. As with three-dimensional spinors, the expression (80.4) gives the rule for forming the scalar product of two spinors.

There is a difference from the three-dimensional case, however, when complex-conjugate spinors are considered. In the theory of three-dimensional spinors (§41), the law of transformation for the complex-conjugate spinor is determined by the requirement that the sum

$$\xi^1\xi^{1*}+\xi^2\xi^{2*}, \tag{80.5}$$

which gives the probability density of the localisation of the particle in space, is a scalar; this led to the relations (41.6) between the coefficients $\alpha, \beta, \gamma, \delta$. In the relativistic theory, the particle density is not a scalar; it is the time component of a four-vector, as already mentioned in §79. Consequently, the above requirement does not apply, and there are no conditions other than (80.3) to be imposed on the coefficients in the transformation (80.2). The four complex quantities related by the single condition (80.3) are equivalent to $8-2 = 6$ real parameters, in accordance with the number of parameters in the Lorentz-group transformations.

Thus the transformation (80.2) and its complex conjugate are essentially different. This means that in the relativistic theory there are spinors of two types. In order to distinguish these, a special notation is used: the indices of a spinor which is transformed by the complex-conjugate formulae to (80.2) are written with dots over them and are called *dotted indices*:

$$\eta = \begin{pmatrix} \eta^{\dot{1}} \\ \eta^{\dot{2}} \end{pmatrix}. \tag{80.6}$$

The rule

$$\eta^{\dot{1}} \sim \xi^{2*}, \quad \eta^{\dot{2}} \sim -\xi^{1*} \tag{80.7}$$

relates the transformation laws for this spinor and for ξ^*; the symbol $\sim$ here and in the rest of §80 denotes "is transformed as".

As already mentioned, the Lorentz group includes, in particular, purely spatial rotations of the three-dimensional coordinates. In respect of these transformations, four-spinors behave in the same way as three-dimensional spinors. The difference between dotted and undotted spinors then disappears, of course, both being transformed in the same manner. (This is the reason for defining the dotted four-spinors by the rule (80.7).) For the complex-conjugate three-dimensional spinor is transformed, as we know (§41), according to $\xi^{1*} \sim \xi^2$, $\xi^{2*} \sim -\xi^1$; comparison with (80.7) thus shows that for spatial rotations

$$\eta^{\dot{1}} \sim \xi^1, \quad \eta^{\dot{2}} \sim \xi^2. \tag{80.8}$$

The four-spinors of higher rank are defined as sets of quantities which are transformed as products of the components of a number of spinors of rank one. The indices of these spinors of higher rank may be partly dotted and partly undotted. For example, there exist three types of spinors of rank two:[†]

$$\xi^{\alpha\beta} \sim \xi^{\alpha}\Xi^{\beta}, \quad \zeta^{\alpha\dot{\beta}} \sim \xi^{\alpha}\eta^{\dot{\beta}}, \quad \eta^{\dot{\alpha}\dot{\beta}} \sim \eta^{\dot{\alpha}}\mathrm{H}^{\dot{\beta}}. \tag{80.9}$$

[†] In §§80–82 the spinor indices, taking the values 1 and 2, will be denoted by the letters at the beginning of the Greek alphabet: $\alpha, \beta, \gamma, \ldots$.

A spinor of rank two has 2×2 = 4 components. If both indices are dotted or both undotted, the spinor can be separated into symmetric and antisymmetric parts: $\frac{1}{2}(\xi^{\alpha\beta}+\xi^{\beta\alpha})$ and $\frac{1}{2}(\xi^{\alpha\beta}-\xi^{\beta\alpha})$. The latter has only one component, $\frac{1}{2}(\xi^{12}-\xi^{21})$, which is a scalar (cf. (80.4)). The symmetrical part is a set of three independent quantities ξ^{11}, ξ^{22}, and $\frac{1}{2}(\xi^{12}+\xi^{21})$, which are transformed into combinations of one another by transformations of the Lorentz group.

For a "mixed" spinor $\zeta^{\alpha\dot{\beta}}$, the order of the indices is arbitrary, since they correspond to different transformation laws. All four components of such a spinor are transformed into combinations of one another, and this number cannot be reduced by taking any linear combinations of the components. A four-vector also has four components, and these are likewise transformed into combinations of one another by Lorentz-group transformations. It is therefore clear that there must be some correspondence between the components of a mixed four-spinor of rank two and those of a four-vector. This correspondence is expressed by the formulae

$$\left.\begin{aligned} \zeta^{1\dot{2}} &= a^3+a^0, & \zeta^{2\dot{1}} &= a^3-a^0, \\ \zeta^{1\dot{1}} &= -a^1+ia^2, & \zeta^{2\dot{2}} &= a^1+ia^2, \end{aligned}\right\} \tag{80.10}$$

where $a^\mu = (a^0, \mathbf{a})$ is a four-vector. The validity of these formulae can be proved as follows.

As already noted, for spatial rotations there is no difference between dotted and undotted spinors, both behaving as three-dimensional spinors. Hence the set of three quantities

$$\zeta^{1\dot{1}} = -a^1+ia^2, \quad \zeta^{2\dot{2}} = a^1+ia^2, \quad \tfrac{1}{2}(\zeta^{1\dot{2}}+\zeta^{2\dot{1}}) = a^3$$

must behave as a three-dimensional symmetrical spinor of rank two, and the above formulae must be the same as the relation established in §41 between the components of such a spinor and those of a three-dimensional vector. A comparison with formulae (41.9) shows that this condition is in fact satisfied.

The antisymmetric combination $\zeta^{1\dot{2}}-\zeta^{2\dot{1}}$ is transformed (under any Lorentz-group transformation) as the difference $\xi^1\eta^{\dot{2}}-\xi^2\eta^{\dot{1}}$; according to the definition (80.7), this implies the correspondence

$$\zeta^{1\dot{2}}-\zeta^{2\dot{1}} \sim \xi^{1}\xi^{1*}+\xi^{2}\xi^{2*}.$$

Such a sum must be the time component of a four-vector, as stated above in connection with (80.5). This condition too is satisfied: according to (80.10) we have

$$\tfrac{1}{2}(\zeta^{1\dot{2}}-\zeta^{2\dot{1}}) = a^{0}.$$

§81. Inversion of spinors

In the exposition (§41) of three-dimensional spinor theory, we have not considered their behaviour under the operation of spatial inversion, since in the non-relativistic theory this would not have led to any new physical results. Here we shall examine the point, however, in order to make clearer the subsequent analysis of the inversion properties of four-spinors.

Inversion is a reversal of the direction of the spatial coordinate axes x, y, z. On repeating the inversion, we return to the original coordinates. For a spinor, however, a return to the original position can be regarded in two different ways, as a rotation through 0° or 360°. These two definitions are not equivalent with respect to spinors, since $\psi = \begin{pmatrix}\psi^{1}\\ \psi^{2}\end{pmatrix}$ changes sign on rotation through 360°. Thus two alternative views of the inversion of spinors are possible: a twofold inversion must either leave a spinor unaltered or change its sign. We shall choose the first of these, and thus suppose that

$$\hat{P}^{2} = +1. \tag{81.1}$$

The choice does not affect the physical results given below.

Inversion of the coordinates changes the sign of polar vectors, but leaves axial vectors unchanged. The latter include angular momentum vectors, and in particular the spin vector. Thus the z-projection of the spin is also unchanged. Hence it follows that, under inversion, each of the components ψ^{1}, ψ^{2} of a three-dimensional spinor, corresponding to a definite value of s_z, can only be transformed into itself. According to the definition (81.1), this means that

$$\hat{P}\psi^{\alpha} = \pm\psi^{\alpha} \qquad (\alpha = 1, 2). \tag{81.2}$$

It must be emphasised, however, that the assignment of a particular parity (+1 or −1) to a spinor has no absolute significance, since spinors change sign on rotation through 2π, and this can always be carried out simultaneously with inversion. The "relative parity" of two spinors ψ and ϕ, defined as the parity of the scalar $\psi^1\phi^2-\phi^2\psi^1$ formed from them, has absolute significance, however; on rotation through 2π, all spinors change sign, and the indeterminacy therefore does not influence the parity (−1 or +1) of this scalar.

Let us now go on to discuss four-dimensional spinors. The requirement that only quantities belonging to the same value of s_z should be transformed into combinations of one another remains valid, of course. But the transformation cannot be simply (81.2) (and a similar one for the dotted spinors); this can be shown, for example, as follows. In consequence of (81.2), the components of four-spinors of higher rank also would be transformed into combinations of one another. But this would contradict formulae (80.10): under inversion of the space coordinates, the components a^1, a^2, a^3 of the (polar) vector **a** change sign, and a^0 remains unchanged; hence $\zeta^{1\dot{2}}$ and $\zeta^{2\dot{1}}$ certainly cannot be transformed into combinations of themselves.

Thus inversion must transform the components of the four-spinor ξ^{α} into expressions involving other quantities. These can only be the components of another spinor $\eta^{\dot{\alpha}}$ whose transformation properties are not the same as those of ξ^{α}. Again treating inversion as an operation satisfying (81.1), we can define its action by

$$\hat{P}\xi^{\alpha} = \eta^{\dot{\alpha}}, \quad \hat{P}\eta^{\dot{\alpha}} = \xi^{\alpha}. \tag{81.3}$$

When this operation is repeated, ξ^{α} and $\eta^{\dot{\alpha}}$ are transformed into themselves, in accordance with the definition (81.1).

Thus the inclusion of inversion as a permissible symmetry transformation requires the simultaneous consideration of the pair of spinors $(\xi^{\alpha}, \eta^{\dot{\alpha}})$, called a *bispinor*.

§82. **Dirac's equation**

A very important case is that of spin $\frac{1}{2}$, which includes the majority of the elementary particles. As will be clear from the foregoing discussion, a wave function that describes such particles in the relativistic theory is a bispinor; it is a set of four components, replacing the two components of the spinor wave function in the non-relativistic theory. Let us construct the wave equation which must be satisfied by the bispinor wave function of a free particle.

From the same arguments as in §79 it is immediately evident that each component of the wave function, when acted on by the operator $\hat{p}_\mu \hat{p}^\mu$, must be multiplied by m^2, i.e. must satisfy the Klein–Fock equation. It is also evident, however, that this equation is here insufficient: of the four components of the bispinor wave function, only two can be linearly independent, in accordance with the number of values that can be taken by the projection of a spin $\frac{1}{2}$. Hence the complete system of wave equations must represent a linear differential relation between the components of the bispinor, obtained by means of the operator $\hat{p}_\mu = i\partial/\partial x^\mu$; this relation must, of course, be expressed by relativistically invariant formulae.

Since the wave function is a set of two spinors (which will be denoted by ξ^α and $\eta^{\dot\alpha}$), in order to obtain the desired result it is reasonable to replace the four-vector $\hat{p}^\mu$ by the equivalent (cf. (80.10)) operator spinor $\hat{p}^{\alpha\dot\beta}$ of rank two, with components

$$\left.\begin{aligned} \hat{p}^{1\dot2} &= \hat{p}^3+\hat{p}^0, & \hat{p}^{2\dot1} &= \hat{p}^3-\hat{p}^0, \\ \hat{p}^{1\dot1} &= -\hat{p}^1+i\hat{p}^2, & \hat{p}^{2\dot2} &= \hat{p}^1+i\hat{p}^2. \end{aligned}\right\} \tag{82.1}$$

We apply the operator $\hat{p}^{\alpha\dot\beta}$ to the spinor ξ^α, forming (in accordance with the rule (80.4)) the scalar product with respect to a pair of undotted indices:

$$\hat{p}^{1\dot\beta}\xi^2-\hat{p}^{2\dot\beta}\xi^1.$$

This product is still a spinor of rank one with respect to the dotted index; it can therefore be expressed only in terms of the dotted spinor $\eta^{\dot\beta}$. Thus we have the equation

$$\hat{p}^{1\dot\beta}\xi^2-\hat{p}^{2\dot\beta}\xi^1 = m\eta^{\dot\beta}, \tag{82.2a}$$

where m is a constant (which will be shown later to be the mass of the particle). Similarly, applying the operator $\hat{p}^{\alpha\dot{\beta}}$ to the spinor $\eta^{\dot{\beta}}$ and forming the scalar product with respect to a pair of dotted indices, we find

$$\hat{p}^{\alpha\dot{2}}\eta^{\dot{1}}-\hat{p}^{\alpha\dot{1}}\eta^{\dot{2}} = m\xi^{\alpha}. \tag{82.2b}$$

The relativistic invariance of these equations is guaranteed by the spinor form in which they are written: the two sides of each equation are either both dotted or both undotted spinors transformed in the same manner under Lorentz transformations.

The relativistic wave equation represented by equations (82.2a, b) is called *Dirac's equation* for a free particle, having been first derived by P. A. M. Dirac in 1928.

Expanding these equations by substituting the expressions (82.1) for the components of the operator $\hat{p}^{\alpha\dot{\beta}}$, we have

$$\left.\begin{aligned}
\hat{p}_0\xi^1-\hat{p}_x\xi^2+i\hat{p}_y\xi^2-\hat{p}_z\xi^1 &= m\eta^{\dot{1}},\\
\hat{p}_0\xi^2-\hat{p}_x\xi^1-i\hat{p}_y\xi^1+\hat{p}_z\xi^2 &= m\eta^{\dot{2}},\\
\hat{p}_0\eta^{\dot{1}}+\hat{p}_x\eta^{\dot{2}}-i\hat{p}_y\eta^{\dot{2}}+\hat{p}_z\eta^{\dot{1}} &= m\xi^1,\\
\hat{p}_0\eta^{\dot{2}}+\hat{p}_x\eta^{\dot{1}}+i\hat{p}_y\eta^{\dot{1}}-\hat{p}_z\eta^{\dot{2}} &= m\xi^2,
\end{aligned}\right\} \tag{82.3}$$

where $\hat{p}_0 = i\partial/\partial t$, and $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$ are the three components of the operator vector $\hat{\mathbf{p}} = -i\nabla$.

For a free particle moving with a definite momentum $\mathbf{p}$ and energy ε, all the components of the wave function are proportional to the factor $e^{i(\mathbf{p}\cdot\mathbf{r}-\varepsilon t)}$ (representing a plane wave). The action of the operator $\hat{p}_0$ multiplies such a function by ε, and that of the operator $\hat{\mathbf{p}}$ multiplies it by $\mathbf{p}$. The differential equations (82.3) are thus reduced to a set of homogeneous linear algebraic equations:

$$\left.\begin{aligned}
(\varepsilon-p_z)\xi^1-(p_x-ip_y)\xi^2 &= m\eta^{\dot{1}},\\
-(p_x+ip_y)\xi^1+(\varepsilon+p_z)\xi^2 &= m\eta^{\dot{2}},\\
(\varepsilon+p_z)\eta^{\dot{1}}+(p_x-ip_y)\eta^{\dot{2}} &= m\xi^1,\\
(p_x+ip_y)\eta^{\dot{1}}+(\varepsilon-p_z)\eta^{\dot{2}} &= m\xi^2,
\end{aligned}\right\} \tag{82.4}$$

Each of the two pairs of equations here determines two components of a bispinor from two other components that are given. If the two pairs of equations are compatible, the result of substituting, for instance, $\eta^{\dot{1}}$ and $\eta^{\dot{2}}$ from the first pair in the second pair must be an identity. It is easily shown that for this to be so we must have

$$\varepsilon^2-p_x^2-p_y^2-p_z^2 = \varepsilon^2-\mathbf{p}^2 = m^2,$$

corresponding to the relativistic expression for the energy of the particle in terms of its momentum, if m is the mass of the particle. This proves the significance of the constant m used in equation (82.2).

The fact that only two of the four components of the bispinor wave function of a free particle can be chosen arbitrarily is in agreement with the fact that, for a given momentum, the state of a particle may still differ as regards the spin projection, which takes two different values.

In the non-relativistic limiting case of small velocities, the particle must be described by a single two-component quantity, a three-dimensional spinor. When the velocity $\mathbf{v}$ tends to zero, so does the momentum $\mathbf{p}$, and the energy ε tends to the rest energy m (in ordinary units, mc^2). From equations (82.4) we then have $\xi^\alpha = \eta^{\dot{\alpha}}$, i.e. the two spinors forming the bispinor in fact become identical.

The two pairs of equations (82.3) can be written more compactly by means of the Pauli matrices defined in §40:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{82.5}$$

If these three matrices are combined into a "matrix vector" $\boldsymbol{\sigma}$, the equations (82.3) can be briefly written as

$$(\hat{p}_0-\hat{\mathbf{p}}\cdot\boldsymbol{\sigma})\,\xi = m\eta, \quad (\hat{p}_0+\hat{\mathbf{p}}\cdot\boldsymbol{\sigma})\,\eta = m\xi. \tag{82.6}$$

When the Pauli matrices are multiplied by the two-component quantities ξ and η, the usual matrix rule is applied, the rows of the matrix being multiplied by the column ξ or η; for example,

$$\sigma_y\xi = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix} = \begin{pmatrix} -i\xi^2 \\ i\xi^1 \end{pmatrix}, \text{ and so on.}$$

§83. Dirac matrices

The spinor form of Dirac's equation is a natural one in the sense that it shows immediately the relativistic invariance of the equation. But when the form of the equation has been established in this way, we can equally well take as the four independent components of the wave function any other linearly independent combinations of the original components. In using Dirac's equation it is in fact usually more convenient to take it in the most general form, where the choice of the wave-function components is not made in advance.

We shall denote the four-component wave function by the symbol Ψ, with components Ψ_i $(i = 1, 2, 3, 4)$; it can be represented as a column[†]

$$\Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \Psi_4 \end{pmatrix}. \tag{83.1}$$

The system of Dirac's equations will be written as

$$\hat{p}_\mu \gamma^\mu_{ik} \Psi_k = m\Psi_i, \tag{83.2}$$

where the γ^μ $(\mu = 0, 1, 2, 3)$ are certain four-rowed matrices with elements γ^μ_{ik} $(i, k = 1, 2, 3, 4)$; the summation on the left of (83.2) is taken both over the matrix (bispinor) index k and over the four-vector index μ.[‡] The matrix indices are usually omitted, the equation being

---

[†] For convenience, the four-component quantity Ψ will be called a bispinor in any representation, and not only in its spinor representation. Correspondingly, the index which labels its components will be referred to as a bispinor index.

[‡] As an example, the following are the matrices γ^μ corresponding to the spinor representation of the wave function. If $\Psi_1 = \xi^1$, $\Psi_2 = \xi^2$, $\Psi_3 = \eta^{\dot{1}}$, $\Psi_4 = \eta^{\dot{2}}$, then

symbolically written

$$(\gamma^{\mu}\hat{p}_{\mu}-m)\,\Psi = 0, \tag{83.3}$$

where

$$\gamma^{\mu}\hat{p}_{\mu} = \hat{p}_0\gamma^0-\hat{\mathbf{p}}\cdot\boldsymbol{\gamma} = i(\gamma^0\,\partial/\partial t+\boldsymbol{\gamma}\cdot\nabla), \tag{83.4}$$

and $\boldsymbol{\gamma}$ denotes a three-dimensional "matrix vector" with components γ^1, γ^2, γ^3. The column notation (83.1) for Ψ corresponds to the fact that the matrices γ^{μ} and Ψ in (83.3) are multiplied according to the ordinary matrix rule: each row of γ^{μ} is multiplied by the column Ψ,

$$(\gamma^{\mu}\Psi)_i = \gamma^{\mu}_{ik}\Psi_k. \tag{83.5}$$

The matrices γ^{μ} are called *Dirac matrices*. In the general case of an arbitrary representation of the wave function, they need only satisfy the conditions which ensure the validity of the equation

$$(\hat{p}^{\mu}\hat{p}_{\mu})\,\Psi = m^2\Psi,$$

i.e. each component of Ψ must satisfy the Klein–Fock equation.

To find these conditions, we multiply equation (83.3) on the left by $\gamma^{\nu}\hat{p}_{\nu}$. Then

$$(\gamma^{\nu}\hat{p}_{\nu})(\gamma^{\mu}\hat{p}_{\mu})\,\Psi = (\gamma^{\nu}\hat{p}_{\nu})\,m\Psi = m^2\Psi.$$

Since all the operators $\hat{p}_{\mu}$ commute, the product $\hat{p}_{\mu}\hat{p}_{\nu}$ is a symmetrical tensor: $\hat{p}_{\mu}\hat{p}_{\nu} = \hat{p}_{\nu}\hat{p}_{\mu}$. The product $\gamma^{\nu}\gamma^{\mu}$ can be separated into a symmetric and an antisymmetric part:

$$\gamma^{\nu}\gamma^{\mu} = \tfrac{1}{2}(\gamma^{\nu}\gamma^{\mu}+\gamma^{\mu}\gamma^{\nu})+\tfrac{1}{2}(\gamma^{\nu}\gamma^{\mu}-\gamma^{\mu}\gamma^{\nu}).$$

$$\gamma^0 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

$$\gamma^2 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \qquad \gamma^3 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

On multiplication by $\hat{p}_\nu \hat{p}_\mu$, the latter part vanishes, leaving

$$\tfrac{1}{2}(\gamma^\nu\gamma^\mu+\gamma^\mu\gamma^\nu)\,\hat{p}_\nu\hat{p}_\mu\Psi = m^2\Psi.$$

In order that the operator on the left should reduce to $\hat{p}_\mu\hat{p}^\mu$, it is necessary that all the pairs of matrices with $\mu \neq \nu$ should anticommute $(\gamma^\mu\gamma^\nu = -\gamma^\nu\gamma^\mu)$, and that the squares of the matrices should be

$$(\gamma^1)^2 = (\gamma^2)^2 = (\gamma^3)^2 = 1, \quad (\gamma^0)^2 = -1 \tag{83.6}$$

(the right-hand sides being taken as unit matrices, of course). All these conditions can be combined in the form

$$\gamma^\mu\gamma^\nu+\gamma^\nu\gamma^\mu = 2g^{\mu\nu}, \tag{83.7}$$

where $g^{\mu\nu}$ is the *metric tensor*, with components

$$g^{\mu\nu} = g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{83.8}$$

The equations (83.7) determine all the properties that are needed in order to use the Dirac matrices. It is not usually necessary to consider the forms of these matrices in any particular representation.

Dirac's equation can be put in a form that is solved for the time derivative, and so a Hamiltonian can be defined for particles with spin ½. Multiplying the equation

$$(\gamma^\mu\hat{p}_\mu - m)\,\Psi = i\gamma^0\,\partial\Psi/\partial t - \boldsymbol{\gamma}\,.\,\hat{\mathbf{p}}\Psi - m\Psi = 0$$

on the left by γ^0, we make the coefficient of $i\partial\Psi/\partial t$ unity (or rather a unit matrix). Thus

$$i\,\partial\Psi/\partial t = (\gamma^0\boldsymbol{\gamma}.\hat{\mathbf{p}}+m\gamma^0)\,\Psi.$$

The operator acting on Ψ on the right-hand side is the Hamiltonian of the particle. It is usually written in the form

$$\hat{H} = \boldsymbol{\alpha}\,.\,\hat{\mathbf{p}}+m\beta, \tag{83.9}$$

with the notation $\boldsymbol{\alpha} = \gamma^0\boldsymbol{\gamma}$, $\beta = \gamma^0$ for the matrices. It is easily seen by means of the relations (83.7) that the square of the operator (83.9) is

$$\hat{H}^2 = \hat{\mathbf{p}}^2 + m^2,$$

as it should be. In this sense we can say that (83.9) is the square root of $\hat{\mathbf{p}}^2 + m^2$!

It has been mentioned at the end of §82 that, in the limiting case of small velocities, the two spinors ξ and η forming the bispinor Ψ are the same. Here, however, we find a certain shortcoming of the spinor form of Dirac's equation: in the limit, all four components of the wave function remain non-zero, although only two of them are really independent. A representation of the wave function in which two components vanish in the limit may therefore be more convenient.

This is achieved by replacing ξ and η by the linear combinations

$$\phi = \frac{1}{\sqrt{2}}(\xi + \eta), \quad \chi = \frac{1}{\sqrt{2}}(\xi - \eta), \tag{83.10}$$

or, in a fuller notation,

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \xi^1 + \eta^{\dot{1}} \\ \xi^2 + \eta^{\dot{2}} \end{pmatrix}, \quad \chi = \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \xi^1 - \eta^{\dot{1}} \\ \xi^2 - \eta^{\dot{2}} \end{pmatrix}.$$

Then $\chi = 0$ for a particle at rest. The representation of Ψ in which its four components are ϕ_1, ϕ_2, χ_1, χ_2 is called the *standard representation.* It will be used in §93 in investigating the motion of an electron in an external field; here, we shall write Dirac's equation in this representation for a free particle. Adding and subtracting equations (82.6) term by term, we have

$$\left.\begin{aligned} \hat{p}_0\phi - \hat{\mathbf{p}} \cdot \boldsymbol{\sigma}\chi &= m\phi, \\ -\hat{p}_0\chi + \mathbf{p} \cdot \boldsymbol{\sigma}\phi &= m\chi. \end{aligned}\right\} \tag{83.11}$$

§84. The current density in Dirac's equation

Let us construct the quantities which act as the particle density ϱ and the particle current density $\mathbf{j}$ in Dirac's equation. In the relativistic theory, these quantities form a four-vector $j^\mu = (\varrho, \mathbf{j})$. They satisfy the equation of continuity, which in four-dimensional form is

$$\partial j^\mu/\partial x^\mu = 0 \tag{84.1}$$

(cf. *Mechanics and Electrodynamics*, §55). This equation expresses the conservation of the quantity

$$Q = \int \varrho \, \mathrm{d}V \tag{84.2}$$

In the non-relativistic theory, this is simply the conservation of the number of particles, but in the relativistic theory (84.1) has a different significance, as will be shown in §86.

The quantities j^μ are expressions bilinear in the wave function Ψ and its complex conjugate Ψ^*. Thus, to find these expressions, we must first determine the form of the equation satisfied by the function Ψ^*. The wave function itself satisfies Dirac's equation:

$$(p_\mu \gamma^\mu - m)\Psi = (i\gamma^0 \, \partial/\partial t + i\boldsymbol{\gamma}\cdot\nabla - m)\Psi = 0. \tag{84.3}$$

The complex conjugate equation is

$$(-i\gamma^{0*} \, \partial/\partial t - i\boldsymbol{\gamma}^*\cdot\nabla - m)\Psi^* = 0.$$

It is seen from the expressions for the matrices γ^0, given in the second footnote to §83, that

$$\gamma^{0+} \equiv \tilde{\gamma}^{0*} = \gamma^0, \quad \boldsymbol{\gamma}^+ = -\boldsymbol{\gamma}, \tag{84.4}$$

i.e. the matrix γ^0 is Hermitian, and γ^1, γ^2, γ^3 are "anti-Hermitian" (the tilde ˜ denotes transposition, i.e. interchange of the rows and columns of the matrix).[†] Hence $\gamma^{0*} = \tilde{\gamma}^0$, $\boldsymbol{\gamma}^* = -\tilde{\boldsymbol{\gamma}}$, and so

$$(-i\tilde{\gamma}^0 \, \partial/\partial t + i\tilde{\boldsymbol{\gamma}}\cdot\nabla - m)\Psi^* = 0.$$

To return to the original (untransposed) matrices, we note that

$$\tilde{\gamma}^\mu \Psi^* \equiv \tilde{\gamma}^\mu_{ik} \Psi^*_k = \Psi^*_k \gamma^\mu_{ki} \equiv \Psi^* \gamma^\mu;$$

in the symbolic notation (without the matrix indices) $\Psi^*\gamma^\mu$, Ψ^* must

---

[†] The expressions given in §83 relate to a specific (spinor) representation of the matrices, but the properties (84.4) are in fact independent of the representation.

be taken as the row

$$\Psi^* = (\Psi_1^*, \Psi_2^*, \Psi_3^*, \Psi_4^*),$$

multiplied by the columns of the matrices γ^μ. Thus we find

$$\Psi^*(-i\gamma^0\,\partial/\partial t + i\boldsymbol{\gamma}\,.\,\nabla - m) = 0,$$

where it is conventionally supposed that the differentiation operators act on the function Ψ^* to their left. Because the first two terms in the parenthesis have opposite signs, they cannot be reduced to a four-dimensional form. To avoid this difficulty, we multiply the whole equation on the right by γ^0, and put $\boldsymbol{\gamma}\gamma^0 = -\gamma^0\boldsymbol{\gamma}$; then

$$\Psi^*\gamma^0(i\gamma^0\,\partial/\partial t + i\boldsymbol{\gamma}\,.\,\nabla + m) = 0.$$

The function $\Psi^*\gamma^0$ is called the *Dirac conjugate* of Ψ, and is denoted by $\overline{\Psi}$:

$$\overline{\Psi} = \Psi^*\gamma^0, \quad \Psi^* = \overline{\Psi}\gamma^0. \tag{84.5}$$

Thus we have, finally,

$$\overline{\Psi}(\hat{p}_\mu\gamma^\mu + m) = 0. \tag{84.6}$$

It is now not difficult to derive an expression for the current density as a four-vector which satisfies the continuity equation (84.1). To do so, we multiply equation (84.6), on the right by Ψ and (84.3) on the left by Ψ^*, and add them term by term. The terms $\pm m\Psi^*\Psi$ cancel, leaving

$$i\frac{\partial\overline{\Psi}}{\partial x^\mu}\gamma^\mu\Psi + i\overline{\Psi}\gamma^\mu\frac{\partial\Psi}{\partial x^\mu} = i\frac{\partial}{\partial x^\mu}(\overline{\Psi}\gamma^\mu\Psi) = 0.$$

This has the form of a continuity equation, in which the current density is represented by the four-vector

$$j^\mu = \overline{\Psi}\gamma^\mu\Psi \tag{84.7}$$

(or, written in full with the matrix indices, $j^\mu = \overline{\Psi}_i\gamma^\mu_{ik}\Psi_k$).

The time component of the four-vector (84.7) is the particle density

$$\begin{aligned}\varrho = \overline{\Psi}\gamma^0\Psi &= \Psi^*\Psi \\ &\equiv |\Psi_1|^2 + |\Psi_2|^2 + |\Psi_3|^2 + |\Psi_4|^2,\end{aligned} \tag{84.8}$$

and the three space components form the three-dimensional current vector

$$\mathbf{j} = \bar{\Psi}\boldsymbol{\gamma}\Psi = \Psi^*\boldsymbol{\alpha}\Psi, \tag{84.9}$$

where $\boldsymbol{\alpha} = \gamma^0\boldsymbol{\gamma}$ is the "matrix vector" already used in (83.9). Here $\boldsymbol{\alpha}$ represents the particle velocity operator.

We can apply (84.7) to the normalisation of a plane wave—the wave function of the state of a free particle with definite values of the momentum $\mathbf{p}$ and the energy ε. For a normalisation to one particle in the volume Ω, we write the wave as

$$\Psi = \frac{1}{\sqrt{\Omega}}u(p)e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})}; \tag{84.10}$$

the wave amplitude $u(p) \equiv u(\varepsilon, \mathbf{p})$ is a constant bispinor depending on the four-momentum of the particle. The components of this bispinor satisfy the algebraic equations

$$(\gamma^\mu p_\mu - m)u = 0, \tag{84.11}$$

which are obtained by substituting (84.10) in Dirac's equation (84.3) (i.e. by simply substituting the quantities p_μ for the operators $\hat{p}_\mu$ in that equation). We shall show that the desired normalisation of the function (84.10) is achieved if the amplitude $u(p)$ is normalised by the condition

$$\bar{u}u = m/\varepsilon. \tag{84.12}$$

Multiplication of equation (84.11) on the left by $\bar{u}$ gives

$$(\bar{u}\gamma^\mu u)p_\mu = m(\bar{u}u) = m^2/\varepsilon,$$

from which we see that $\bar{u}\gamma^\mu u = p^\mu/\varepsilon$, and the current four-vector is therefore

$$j^\mu = \bar{\Psi}\gamma^\mu\Psi = \frac{1}{\Omega}\bar{u}\gamma^\mu u = \frac{p^\mu}{\Omega\varepsilon}. \tag{84.13}$$

The particle density $\varrho = p^0/\varepsilon\Omega = 1/\Omega$, in accordance with the required normalisation. The three-dimensional current density is $\mathbf{j} = \mathbf{p}/\varepsilon\Omega = \mathbf{v}/\Omega$, where $\mathbf{v}$ is the velocity of the particles.

CHAPTER 13

PARTICLES AND ANTIPARTICLES

§85. Ψ-operators

It has been shown in Chapter 11 how a quantum description of a free electromagnetic field can be constructed, starting from known properties of the field in the classical limit and using the ideas of ordinary quantum mechanics. The description thus obtained, in which the field is regarded as a system of photons, has many features that are applicable also to the relativistic description of particles in quantum theory.

The electromagnetic field is a system with an infinite number of degrees of freedom. It is not subject to a law of conservation of number of particles (photons), and its possible states include some in which the number of particles is arbitrary.[†] But this must be a general property of systems of any particles in the relativistic theory. The conservation of particle number in the non-relativistic theory is connected with the law of conservation of mass: the sum of the (rest) masses of the particles is unaltered by their interaction, and the constancy of the sum of the masses in a system of particles implies that their number is also unchanged. In the relativistic theory, however, there is no conservation of mass; only the total energy of the system (including the rest energy of the particles) is conserved. Thus the number of particles need not be conserved, and therefore any relativistic theory of particles must be a theory with an infinite number of degrees

---

[†] In practice, of course, the number of photons changes only as a result of various interaction processes.

of freedom. That is to say, such a theory of particles has the character of a field theory.

An adequate mathematical formalism for the description of systems having a variable number of particles is that of second quantisation, in which the independent variables are the occupation numbers of the various states of a particle. In the quantum description of the electromagnetic field, the field potential $\hat{\mathbf{A}}$ appears as a second-quantisation operator. It can be expressed in terms of the wave functions of individual photons and their creation and annihilation operators. The quantised wave function operator plays a similar part in the description of a system of particles.

The arguments given in the present section apply equally to particles with any spin. We shall therefore not specify the mathematical nature of the wave functions. For example, the plane waves will be written as

$$\Psi_{\mathbf{p}} = \frac{1}{\sqrt{\Omega}} u(p) e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})}, \tag{85.1}$$

with the assumption that the wave amplitude $u(p)$, a function of the four-momentum, may be a scalar (for spin-zero particles), a bispinor (for spin-$\frac{1}{2}$ particles), and so on.

In accordance with the general procedure of the second quantisation method, we have to consider the expansion of an arbitrary wave function in terms of the eigenfunctions of a complete set of possible states of a free particle, the plane waves $\Psi_{\mathbf{p}}$:[†]

$$\Psi = \sum_{\mathbf{p}} a_{\mathbf{p}} \Psi_{\mathbf{p}}, \qquad \Psi^* = \sum_{\mathbf{p}} a_{\mathbf{p}}^* \Psi_{\mathbf{p}}^*.$$

The coefficients $a_{\mathbf{p}}$, $a_{\mathbf{p}}^*$ are then to be regarded as the operators $\hat{a}_{\mathbf{p}}$, $\hat{a}_{\mathbf{p}}^+$ for annihilation and creation of particles in the corresponding states.

Here, however, we immediately encounter a difference of principle as compared with the non-relativistic theory. If the plane wave (85.1) satisfies the wave equation, only the condition $\varepsilon^2 = \mathbf{p}^2 + m^2$ need be satisfied; the energy itself can have two values, $\varepsilon = \pm\sqrt{(\mathbf{p}^2 + m^2)}$.

[†] For particles with spin, the summation must also be taken over the polarisations of the particle; the corresponding index will be omitted, for brevity.

Only positive values of ε can have the physical significance of the energy of a free particle. But the negative values cannot be simply omitted: the general solution of the wave equation can be obtained only by superposing all its independent particular solutions. This shows that the interpretation of the expansion coefficients for Ψ and Ψ^* in the second quantisation method must be somewhat different.

We may write the expansion in the form

$$\Psi = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{p}} a_{\mathbf{p}}^{(+)} u(\varepsilon, \mathbf{p}) e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})} + \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{p}} a_{\mathbf{p}}^{(-)} u(-\varepsilon, \mathbf{p}) e^{i(\varepsilon t + \mathbf{p}\cdot\mathbf{r})}, \tag{85.2}$$

where the first sum contains plane waves with positive "frequency" and the second sum those with negative "frequency"; ε always denotes the positive quantity $+\sqrt{(\mathbf{p}^2 + m^2)}$. In the second quantisation, the coefficients $a_{\mathbf{p}}^{(+)}$ in the first sum are replaced as usual by the particle annihilation operators $\hat{a}_{\mathbf{p}}$.

In the second sum, we first of all replace the summation variable $\mathbf{p}$ by $-\mathbf{p}$; since the summation is over all possible values of $\mathbf{p}$, this of course does not affect either the range of summation or the value of the sum. After the change, the exponential factor in each term becomes $e^{i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})}$, which is the same as for the complex conjugate wave functions Ψ^* with "positive" frequencies. Such functions are to be multiplied in the second quantisation by the particle creation operators. Accordingly, we replace the coefficients $a_{-\mathbf{p}}^{(-)}$ by the creation operators $\hat{b}_{\mathbf{p}}^{+}$ for some other particles, which are in general different from those to which the operators $\hat{a}_{\mathbf{p}}$ relate. We thus obtain the Ψ-operators in the form

$$\left.\begin{aligned} \hat{\Psi} &= \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{p}} \left\{\hat{a}_{\mathbf{p}} u(p) e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})} + \hat{b}_{\mathbf{p}}^{+} u(-p) e^{i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})}\right\}, \\ \hat{\Psi}^{+} &= \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{p}} \left\{\hat{a}_{\mathbf{p}}^{+} u^*(p) e^{i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})} + \hat{b}_{\mathbf{p}} u^*(-p) e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})}\right\}, \end{aligned}\right\} \tag{85.3}$$

with the notation $u(-p) \equiv u(-\varepsilon, -\mathbf{p})$.

Thus all the operators $\hat{a}_{\mathbf{p}}$ and $\hat{b}_{\mathbf{p}}$ are multiplied by functions having the "correct" time dependence ($\sim e^{-i\varepsilon t}$), while the operators $\hat{a}_{\mathbf{p}}^{+}$ and

$\hat{b}_\mathbf{p}^+$ are multiplied by the complex conjugate functions. This makes it possible to interpret the former operators, in accordance with the general rules, as annihilation operators for particles with momentum **p** and energy ε, and the latter as creation operators for these particles.

In this way we arrive at the conception of particles of two types which occur simultaneously and on an equal footing. These are called *particles* and *antiparticles*; the significance of the names will be shown in §86. One type corresponds to the operators $\hat{a}_\mathbf{p}$, $\hat{a}_\mathbf{p}^+$ in the second quantisation formalism, and the other type to $\hat{b}_\mathbf{p}$, $\hat{b}_\mathbf{p}^+$. The two types of particle have the same mass, since their operators appear in the same Ψ-operator, which satisfies the same wave equation.

§86. Particles and antiparticles

In order to elucidate further the properties and interrelation of particles and antiparticles, we must derive expressions for the operators of the total energy and the total number of particles in the system. The derivation depends on the spin of the particles; let us consider a field of particles with spin $\frac{1}{2}$ (a *spinor field*).

To derive the desired expressions, it is then sufficient to know that for particles described by Dirac's equation there exists a Hamiltonian, and that the particle density is represented by the product $\Psi^*\Psi$. These facts enable us to make direct use of the results obtained in the non-relativistic theory in §§47 and 48 (where both the properties mentioned above occur for particles with any spin).[†]

We have seen that, in the mathematical formalism of second quantisation, the Hamiltonian $\hat{H}$ of a system of particles is found from the Hamiltonian $\hat{H}^{(1)}$ of a single particle as the integral[‡]

$$\hat{H} = \int \hat{\Psi}^+ \hat{H}^{(1)} \hat{\Psi}\, \mathrm{d}V. \tag{86.1}$$

In the non-relativistic theory, this gave a trivial result: substitution

---

[†] It should also be recalled that *neither* property occurs for relativistic particles with spin zero, described by the scalar Klein–Fock equation (§79).

[‡] The index (1) in the single-particle Hamiltonian is used here to distinguish it from that of the whole system.

of the Ψ-operators

$$\hat{\Psi} = \sum_{\mathbf{p}} \hat{a}_{\mathbf{p}}\Psi_{\mathbf{p}}, \quad \hat{\Psi}^{+} = \sum_{\mathbf{p}} \hat{a}_{\mathbf{p}}^{+}\Psi_{\mathbf{p}}^{*}, \tag{86.2}$$

gave, whatever the commutation rules for the operators $\hat{a}_{\mathbf{p}}$, $\hat{a}_{\mathbf{p}}^{+}$,

$$\hat{H} = \sum_{\mathbf{p}} \varepsilon_{\mathbf{p}}\hat{a}_{\mathbf{p}}^{+}\hat{a}_{\mathbf{p}}, \tag{86.3}$$

where $\varepsilon_{\mathbf{p}}$ are the eigenvalues of the Hamiltonian $\hat{H}^{(1)}$, i.e. the free-particle energies. The eigenvalues of the operator products $\hat{a}_{\mathbf{p}}^{+}\hat{a}_{\mathbf{p}}$ are the occupation numbers $N_{\mathbf{p}}$ of the states; the eigenvalues of the total energy of the system were therefore equal to the expression $E = \Sigma\varepsilon_{\mathbf{p}}N_{\mathbf{p}}$, which is obvious.

Similarly, a trivial result was obtained for the total number of particles in the system, whose operator is given by the integral

$$\hat{N} = \int \hat{\Psi}^{+}\hat{\Psi}\,\mathrm{d}V. \tag{86.4}$$

Substitution of the Ψ-operators (86.2) gave

$$\hat{N} = \sum_{\mathbf{p}} \hat{a}_{\mathbf{p}}^{+}\hat{a}_{\mathbf{p}}, \tag{86.5}$$

so that the eigenvalues were $N = \Sigma N_{\mathbf{p}}$.

In the relativistic theory, the existence of negative eigenvalues of the particle Hamiltonian $\hat{H}^{(1)}$ entirely changes the situation. Instead of (86.3), we now have

$$\hat{H} = \sum_{\mathbf{p}} \varepsilon_{\mathbf{p}}\hat{a}_{\mathbf{p}}^{+}\hat{a}_{\mathbf{p}} - \sum_{\mathbf{p}} \varepsilon_{\mathbf{p}}\hat{b}_{\mathbf{p}}\hat{b}_{\mathbf{p}}^{+}. \tag{86.6}$$

The first sum corresponds to positive eigenvalues $\varepsilon_{\mathbf{p}} = +\sqrt{(\mathbf{p}^2+m^2)}$, and has the same form as (86.3). The second sum corresponds to negative eigenvalues $-\varepsilon_{\mathbf{p}}$; this accounts for the negative sign of the sum. The order of the factors $\hat{b}_{\mathbf{p}}$ and $\hat{b}_{\mathbf{p}}^{+}$ in the second sum differs from that in the first, because the Ψ-operators (85.3) have $\hat{b}_{\mathbf{p}}^{+}$ and $\hat{b}_{\mathbf{p}}$ coupled with $\hat{a}_{\mathbf{p}}$ and $\hat{a}_{\mathbf{p}}^{+}$ respectively. Similarly, for the operator (86.4), which will now be denoted by $\hat{Q}$, we have instead of (86.5)

$$\hat{Q} = \sum_{\mathbf{p}} \hat{a}_{\mathbf{p}}^{+}\hat{a}_{\mathbf{p}} + \sum_{\mathbf{p}} \hat{b}_{\mathbf{p}}\hat{b}_{\mathbf{p}}^{+}. \tag{86.7}$$

To determine the eigenvalues of the operators (86.6) and (86.7), we must first put the factors in the second sums in the order $\hat{b}_{\mathrm{p}}^{+}\hat{b}_{\mathrm{p}}$, for which the eigenvalues are equal to the occupation numbers. Here, however, the commutation rules obeyed by the particle creation and annihilation operators become important.

It is easy to see that a reasonable result for the eigenvalues of the Hamiltonian (86.6) is obtained only if these operators satisfy the Fermi commutation rules:

$$\left.\begin{aligned} \hat{a}_{\mathrm{p}}\hat{a}_{\mathrm{p}}^{+}+\hat{a}_{\mathrm{p}}^{+}\hat{a}_{\mathrm{p}} &= 1, \\ \hat{b}_{\mathrm{p}}\hat{b}_{\mathrm{p}}^{+}+\hat{b}_{\mathrm{p}}^{+}\hat{b}_{\mathrm{p}} &= 1: \end{aligned}\right\} \tag{86.8}$$

in this case, the Hamiltonian (86.6) becomes

$$\hat{H} = \sum_{\mathrm{p}} \varepsilon_{\mathrm{p}}(\hat{a}_{\mathrm{p}}^{+}\hat{a}_{\mathrm{p}}+\hat{b}_{\mathrm{p}}^{+}\hat{b}_{\mathrm{p}}-1).$$

The eigenvalues of the products $\hat{a}_{\mathrm{p}}^{+}\hat{a}_{\mathrm{p}}$ and $\hat{b}_{\mathrm{p}}^{+}\hat{b}_{\mathrm{p}}$ are positive integers N_{p} and $\bar{N}_{\mathrm{p}}$, the numbers of particles and antiparticles in the corresponding states. The infinite additive constant $-\Sigma\varepsilon_{\mathrm{p}}$, the "energy of the vacuum", may simply be omitted, as was done for a similar reason in the case of photons (§77). Then the energy of the system is given by the essentially positive expression

$$E = \sum_{\mathrm{p}} \varepsilon_{\mathrm{p}}(N_{\mathrm{p}}+\bar{N}_{\mathrm{p}}), \tag{86.9}$$

in accordance with the idea that two kinds of particle actually exist: the total energy of the system is equal to the sum of the energies of all the particles and antiparticles in it.

If, instead of (86.8), we used the Bose commutation rules (commutators instead of anticommutators), we should obtain

$$\hat{H} = \sum_{\mathrm{p}} \varepsilon_{\mathrm{p}}(\hat{a}_{\mathrm{p}}^{+}\hat{a}_{\mathrm{p}}-\hat{b}_{\mathrm{p}}^{+}\hat{b}_{\mathrm{p}}+1),$$

and instead of (86.9) the physically meaningless expression $\Sigma\varepsilon_{\mathrm{p}}(N_{\mathrm{p}}-\bar{N}_{\mathrm{p}})$, which is not positive-definite and hence cannot represent the energy of a system of free particles.

Having thus established the commutation rules for the particle annihilation and creation operators, let us now consider the operator (86.7). Using (86.8) to change the order of factors in the second sum, we find

$$\hat{Q} = \sum_{\mathbf{p}} (\hat{a}_{\mathbf{p}}^{+}\hat{a}_{\mathbf{p}} - \hat{b}_{\mathbf{p}}^{+}\hat{b}_{\mathbf{p}} + 1).$$

The eigenvalues of this operator are (again omitting the unimportant additive constant $\Sigma 1$)

$$Q = \sum_{\mathbf{p}} (N_{\mathbf{p}} - \bar{N}_{\mathbf{p}}), \qquad (86.10)$$

and are therefore equal to the differences between the total numbers of particles and antiparticles.

This is a very important result. The operator $\hat{Q}$ corresponds to the quantity (84.2) whose conservation is expressed by the equation of continuity (84.1). We now see that this conservation law does not imply the conservation of the numbers of particles and antiparticles separately, nor of their sum. Only the difference of these numbers must be conserved. In other words, only particle–antiparticle pairs can be formed or disappear in various interactions.[†] All such processes must, naturally, conserve the energy and momentum of the whole system of interacting particles. In particular, the disappearance of a pair in a particle–antiparticle collision must be accompanied by the appearance of some other particles so as to ensure the conservation of energy and momentum; these may be photons, in which case the process is called *pair annihilation*.

If a particle is electrically charged, its antiparticle must have a charge of the opposite sign, for, if they had like charges, the appearance or disappearance of the pair would contravene a rigorous law of nature, the conservation of total electric charge.

The quantity Q is sometimes called the *field charge* of the particles concerned. For electrically charged particles, Q determines the total

---

[†] Here it is, of course, assumed that the interaction does not violate the conservation of Q. This assumption is valid for all interactions that are known to occur.

charge of the system in terms of the unit charge e. But particles and antiparticles may also be electrically neutral.[†]

Thus we see that the nature of the relativistic relation between the energy and the momentum (the twofold root of the equation $\varepsilon^2 = \mathbf{p}^2 + m^2$), together with the requirements of relativistic invariance, leads in the quantum theory to a new principle of classification of particles: there can exist pairs of different particles (particle and antiparticle) which are interrelated in the way described above. This remarkable prediction was first made by Dirac in 1930, before the discovery of the first antiparticle, the positron or antielectron.[‡]

§87. The relation between the spin and the statistics

The results described in §86 have another important aspect. We have seen that the natural requirements of physics necessarily have the result that spin-$\frac{1}{2}$ particles obey Fermi statistics. Hence in turn there follows the general conclusion that all particles with half-integral spin are fermions, while those with integral (including zero) spin are bosons.[||]

This is obvious if we note that, as regards its spin properties, any particle with non-zero spin s can be regarded as "composed" of $2s$ particles with parallel spins of $\frac{1}{2}$ (and a particle with spin zero can be regarded as two particles with antiparallel spins of $\frac{1}{2}$). When s is half-integral, the number $2s$ is odd; when s is integral, $2s$ is even. A "complex" particle formed from an odd number of fermions is itself a fermion, and one formed from an even number of fermions is a boson,

---

[†] The neutrons and the neutrinos (spin $\frac{1}{2}$) are neutral fermions. The neutral kaons (spin 0) are neutral bosons.

[‡] Dirac himself arrived at the idea of a positron as a "hole" in a continuum of occupied electron states of negative energy. This concept, however, obviously cannot be taken literally, and, moreover, it is inadequate in the sense that the notion of particles and antiparticles actually applies to particles with any spin, not only those with half-integral spin, for which Pauli's principle is valid.

[||] The integral-spin particles include photons. The fact that photons are bosons has already been demonstrated in §77 from the analogy with oscillators, i.e. essentially from the properties of the electromagnetic field in the classical limit.

as already discussed in §45: the statistics is decided by the behaviour of the wave function of the system of particles when any pair of them are interchanged, the wave function changing sign when fermions are interchanged but not when bosons are interchanged. The interchange of two particles with half-integral spin is equivalent, from the above discussion, to a simultaneous interchange of an odd number of pairs of fermions with spin $\frac{1}{2}$, and therefore changes the sign of the wave function. The interchange of two particles with integral spin is equivalent to an interchange of an even number of pairs of fermions, and therefore does not change the sign of the wave function.

The specific features of spin-$\frac{1}{2}$ particles used in the analysis in §86 were only the existence of a Hamiltonian and the expression $\Psi^*\Psi$ for the particle density. Both of these are due to the spinor properties of the wave functions of such particles and to the properties of Dirac's equation, which is satisfied by these functions. In turn, all these properties are essentially a consequence of just the conditions of relativistic invariance and the isotropy of space (i.e. a consequence of the symmetry under Lorentz-group transformations). In this sense, we can say that the relation between the spin and the statistics obeyed by the particles is also a direct consequence of these conditions.[†] The origin of this relation was first elucidated by W. Pauli (1940).

§88. **Strictly neutral particles**

In the second quantisation of the wave function (85.2), the coefficients $a_{\mathrm{p}}^{(+)}$ and $a_{\mathrm{p}}^{(-)}$ were replaced by the annihilation and creation operators for different particles. This is not necessary, however: as a particular case, the annihilation and creation operators in $\hat{\Psi}$ may relate to the same particles. All that is necessary is that the annihilation

---

[†] The generalisation of the relation between the spin and the statistics from the case of spin $\frac{1}{2}$ to particles with any spin has been based here on the consideration of "composite" particles, but a similar result could also be reached by considering the mathematical structure of the expressions corresponding to the operators $\hat{H}$ and $\hat{Q}$ for the fields of these particles, constructed in accordance with the requirements of relativistic invariance.

operators should be in the "positive-frequency" wave functions, and the creation operators in the "negative-frequency" wave functions. Then, denoting these operators by $\hat{c}_{\mathbf{p}}$ and $\hat{c}_{\mathbf{p}}^{+}$,

$$\hat{\Psi} = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{p}} \left\{ \hat{c}_{\mathbf{p}} u(p) e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})} + \hat{c}_{\mathbf{p}}^{+} u^{*}(-p) e^{i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})} \right\}. \tag{88.1}$$

The field described by the Ψ-operator corresponds to a system of particles of one kind only, which may be said to be their own antiparticles.

It is evident that the electric charge of such particles must certainly be zero. They are said to be *strictly neutral*, as opposed to electrically neutral particles which are not their own antiparticles.

For strictly neutral particles there is no law of conservation of the field "charge" Q: the identity of particle and antiparticle corresponds to identical equality of the numbers $N_{\mathbf{p}}$ and $\bar{N}_{\mathbf{p}}$, so that the quantity (86.10) is identically zero. Because this limitation is removed, strictly neutral particles can be created or annihilated (into photos) singly, and not necessarily in pairs.

Among the "elementary" particles with spin zero, the neutral pions are strictly neutral. An example of a strictly neutral "composite" particle is *positronium*, a hydrogen-like system consisting of a positron and an electron; the spin of positronium may be 0 or 1. No strictly neutral particles with half-integral spin are known.

The structure of the Ψ-operator (88.1) is similar to that of the electromagnetic field operator (76.15): in both, the particle annihilation and creation operators appear in the same field operator. In this sense one can say that the photons themselves are strictly neutral particles. Their creation or annihilation is the familiar emission or absorption of photons by a system of charged particles.

The existence of a new symmetry property causes the particle to have a new property with no analogue in the non-relativistic theory. This concerns the transformation of *charge conjugation*, i.e. the interchange of particle and antiparticle, whose operator is denoted by $\hat{C}$. If a particle (or a system of particles) is not strictly neutral, charge conjugation involves its replacement by a different physical system, for instance the replacement of a system of electrons by a system of positrons; this does not lead to any new property of the particle itself. But if the particle (or system) is strictly neutral, charge conjugation leaves it unchanged. We can therefore consider the behaviour of the wave function under this transformation, and hence the eigenvalues of the operator $\hat{C}$. A twofold application of charge conjugation is, of course, an identity transformation: $\hat{C}^2 = 1$. Like any operator having this property, it has the eigenvalues $C = \pm 1$; these are called the *charge parity*. If a system has a definite charge parity, its wave functions are unchanged or change sign under charge conjugation, the system being said to be charge-even or charge-odd respectively.

As an example, let us determine the charge parity of positronium (see above). To describe the charge symmetry of the system, we must regard the particle and antiparticle (in this case the electron and the positron) as two different "charge states" of one particle, differing in the value of the "charge quantum number" $Q = \pm 1$. The wave function of the system is represented as the product of an orbital factor (depending on the coordinates of the particles), a spin factor, and a "charge" factor: $\Psi = \Psi_{\text{orb}} \Psi_{\text{spin}} \Psi_{\text{ch}}$.

In this case, charge conjugation is equivalent to an interchange of the two particles. Interchanging the coordinates of the two particles is, in turn, equivalent to inversion about the mid-point of the line joining the particles; it multiplies Ψ_{orb} by $(-1)^l$, where l is the orbital angular momentum of the positronium (see (19.5)). The spin function is symmetrical with respect to interchange of the particles if their spins are parallel (total spin $S = 1$) and antisymmetrical if the spins are antiparallel ($S = 0$); see §46. Thus Ψ_{spin} is multiplied by $(-1)^{S+1}$. Finally, Ψ_{ch} is multiplied by the required value of C.

The interchange of two fermions must change the sign of the complete wave function Ψ. Thus we must have $(-1)^l(-1)^{S+1}C = -1$, whence

$$C = (-1)^{l+S}. \tag{88.2}$$

The levels with spin $S = 0$ are called *parapositronium* levels, and those with $S = 1$ *orthopositronium* levels. In the ground state $l = 0$, so that the ground states of parapositronium and orthopositronium are respectively charge-even ($C = 1$) and charge-odd ($C = -1$).

Positronium is unstable, the electron and the positron ultimately undergoing mutual annihilation. The charge parity of positronium places certain restrictions on the ways in which this annihilation can occur. We shall see later that the photon is a charge-odd particle (cf. the first footnote to §95). Hence, for example, in the ground state of parapositronium ($C = 1$), annihilation can occur with the formation of two photons (the charge parity of the two-photon system being $C = -1 \times -1 = 1$), but in the ground state of orthopositronium ($C = -1$) decay into two photons is impossible, and the annihilation takes place with the formation of three photons.[†]

The neutral pion mentioned above is also unstable and decays into two photons. Hence it follows that this particle is charge-even; its decay into an odd number of photons is forbidden for that reason.[‡]

§89. **Internal parity of particles**

We have already seen, in the discussion of the non-relativistic quantum theory, how the symmetry with respect to inversion of the space coordinates leads to the occurrence of a new characteristic of the state of a particle, namely its parity. The relativistic theory adds another aspect to this concept.

Let us first consider particles with spin zero, described by scalar wave functions. Scalars can be of two kinds, the difference between which lies precisely in their behaviour under inversion. Inversion changes the sign of the coordinates in the arguments of the function, and may also either change or not change the sign of the function itself:

$$\hat{P}\Psi(t, \mathbf{r}) = \pm\, \Psi(t, -\mathbf{r}), \tag{89.1}$$

the signs + and − on the right corresponding to true scalars and pseudoscalars respectively.

[†] The lifetime of parapositronium (i.e. the reciprocal of its decay probability) is 1.2×10^{-10} sec. That of orthopositronium is much longer (1.4×10^{-7} sec), because of the smaller probability of decay into a greater number of photons.

[‡] In this argument it is tacitly assumed that the charge parity of the system is conserved. We shall return to this point in §90.

Hence we see that two features of the behaviour of the wave function under inversion must be distinguished. One of these relates to the dependence of the wave function on the coordinates. In non-relativistic quantum mechanics, only this aspect was considered; it leads to the concept of the parity of a state (which we shall here call the *orbital parity*), describing the symmetry properties of the motion of the particle. If the state has a definite orbital parity (+1 or −1), this means that

$$\Psi(t, -\mathbf{r}) = \pm\Psi(t, \mathbf{r}).$$

The other feature is the behaviour of the wave function at a given point (which may conveniently be taken as the origin) under inversion of the coordinate axes. This leads to the concept of the *internal parity* of the particle. The two signs in the definition (89.1) correspond to internal parity +1 and −1 (for a particle with spin zero). The total parity of a system of particles is given by the product of their internal parities and the orbital parity of their relative motion.

The "internal" symmetry properties of various particles appear, of course, only in their mutual transformation processes. In non-relativistic quantum mechanics, the analogue of the internal parity is the parity of a bound state of a composite system, such as a nucleus. In the relativistic theory, which makes no essential difference between composite and elementary particles, this internal parity is no different from the internal parity of those particles which are regarded as elementary in the non-relativistic theory. In the non-relativistic case, where these particles are to be regarded as unalterable, their internal symmetry properties are not observable, and a discussion of these would be devoid of physical significance.

The concept of internal parity can be readily formulated in the rest frame of the particle, where the wave function reduces to a quantity independent of the coordinates (the wave amplitude u in the functions (85.1)). For particles with spin zero, this quantity is a scalar or a pseudoscalar, the transformation of which amounts simply to a multiplication by +1 or −1.

For a particle with spin $\frac{1}{2}$, the wave function in the rest frame reduces to a single three-dimensional spinor (see the end of §82). The concept of the internal parity of the particle depends on the behaviour of this spinor under inversion. It has already been mentioned in §81 that, although the two possible transformation laws for three-dimensional spinors (the two signs in (81.2)) are not mutually equivalent, the assignment of a particular parity to a spinor has no absolute significance. There is therefore also no meaning in speaking of the internal parity of a spin-$\frac{1}{2}$ particle by itself. We can, however, refer to the relative internal parity of two such particles.

Let us consider from this standpoint the relative internal parity of a particle and its antiparticle. For a spin-zero particle, the question is trivial: such particles and antiparticles are described by the same (scalar or pseudoscalar) wave functions, and their internal parities are therefore obviously the same.

Two spinors $\xi = \begin{pmatrix} \xi^1 \\ \xi^2 \end{pmatrix}$ and $\eta = \begin{pmatrix} \eta^{\dot{1}} \\ \eta^{\dot{2}} \end{pmatrix}$ forming a bispinor $\Psi = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$ describing a particle with spin $\frac{1}{2}$ (say, an electron) reduce to the same three-dimensional spinor in the rest frame of the particle, which we denote by $\Phi^{(e)} = \begin{pmatrix} \Phi^1 \\ \Phi^2 \end{pmatrix}$:

$$\xi = \eta = \Phi^{(e)}. \tag{89.2}$$

The inversion operation defined according to (81.3) replaces ξ by η; from (89.2), this definition corresponds to a transformation of the three-dimensional spinor $\Phi^{(e)}$ such that

$$\hat{P}\Phi^{(e)} = \Phi^{(e)}. \tag{89.3}$$

The positron corresponds to the "negative-frequency" wave functions arising from Dirac's equation with the sign of the four-momentum p^μ changed (in the Ψ-operators (85.3), the positron operators $\hat{b}_{\mathrm{p}}, \hat{b}_{\mathrm{p}}^+$ appear as coefficients of wave functions with amplitudes $u(-p)$). The equation (89.2) for an electron in the rest frame followed from Dirac's equations (82.4) with $\mathbf{p} = 0$, $\varepsilon = m$. If $(\varepsilon, \mathbf{p})$ in these equations is replaced by $(-\varepsilon, -\mathbf{p})$ and we then put $\mathbf{p} = 0$, $\varepsilon = m$, the result is

$$\xi = -\eta \equiv \Phi^{(p)}. \tag{89.4}$$

The inversion operation, replacing ξ by η, will now correspond to the following transformation of the three-dimensional spinor $\Phi^{(p)}$:

$$\hat{P}\Phi^{(p)} = -\Phi^{(p)}, \tag{89.5}$$

with the opposite sign to that in (89.3). Hence a scalar constructed from products of components of $\Phi^{(e)}$ and $\Phi^{(p)}$ will change sign on inversion. We therefore conclude that the internal parities of particle and antiparticle with spin $\frac{1}{2}$ are opposite (V. B. Berestetskiĭ 1948).

§90. The *CPT* theorem

The space–time symmetry properties of physical phenomena are expressed by the invariance of the equations describing them, under various transformations of the four-dimensional coordinate system.

A universal law of nature is that of relativistic invariance, i.e. invariance under transformations of the Lorentz group.[†] As shown in §80, these include both the ordinary three-dimensional rotations and the Lorentz transformations (rotations of the four-dimensional coordinate system, which change the direction of the time axis).

As well as these transformations, there are others which do not reduce to rotations: spatial inversion (reversal of the direction of the three spatial axes) and time reversal (reversal of the direction of the time axis). The invariance under spatial inversion (P invariance) expresses the mirror symmetry of space. The invariance under time reversal (T invariance) expresses the equivalence of the two directions of time. Both these are valid for phenomena described by the non-relativistic theory.

For phenomena pertaining to the relativistic case, however, the symmetry under spatial inversion (and the related law of conservation of spatial parity) is no longer universal. The available experimental data show that this symmetry is conserved in electromagnetic interactions and in what are called *strong interactions* (nuclear forces).

[†] To avoid misunderstanding it should be emphasised that the phenomena under consideration do not involve gravitational fields.

It is, however, violated in *weak interactions*, i.e. those which are responsible for the majority of slow decays of elementary particles (e.g. β-decay).[†]

In weak interactions the symmetry between particles and antiparticles expressed by the transformation of charge conjugation (C invariance) is also violated. There are, however, no experimental results indicating a violation of this symmetry in electromagnetic and strong interactions.

The loss of symmetry under spatial inversion in certain interaction processes need not imply that space has only mirror symmetry. The symmetry of space could be "saved" if there were a universal law of nature asserting invariance under a transformation consisting of simultaneous inversion and charge conjugation (the CP transformation, or *combined inversion*).[‡] In this transformation, particles are replaced by antiparticles simultaneously with spatial inversion. If there is CP invariance, the processes involving particles and those involving antiparticles would differ by spatial inversion. On this view, space would remain entirely symmetrical, the asymmetry being transferred to the charged particles. This asymmetry would not affect the symmetry of space, in the same way as the latter is not affected by the existence of stereoisomeric molecules (those which are related in the same way as an object and its mirror image).

These arguments are, however, not entirely confirmed by experiment. Although the majority of weak-interaction processes are in fact CP-invariant, there are some which are not. It is as yet unclear what will be the significance of such violations of CP invariance in future theories.

Thus the requirements of symmetry with respect to each of the transformations C and P (and T) separately are not universal laws of nature. Their universality, it should be emphasised, is not a logical consequence of the principles of the existing theory, as well as not being confirmed by experiment. These principles do, however, imply

---

[†] The idea that parity might not be conserved in weak interactions was first put forward by T. D. Lee and C. N. Yang (1956).

[‡] These ideas were put forward by L. D. Landau (1957).

invariance under the three transformations simultaneously. We shall show how this symmetry arises as a natural consequence of the requirements of relativistic invariance.

In order to clarify the argument, let us first recall some concepts pertaining to transformations of three-dimensional space.

A reversal of the direction of one of the coordinate axes x, y, z is a mirror reflection in a certain plane; for example, the transformation $x \to -x$, $y \to y$, $z \to z$ is a reflection in the yz-plane. This transformation cannot be reduced to any rotations of the coordinates. A reversal of the directions of two axes, however, is equivalent to a rotation; for example, the transformation $x \to -x$, $y \to -y$, $z \to z$ is a rotation through 180° about the z-axis. Finally, a reversal of the directions of all three axes (inversion of the coordinates) is a transformation that cannot be reduced to rotations; inversion and reflection in a plane are, however, reducible to one another, in the sense that they differ only by a rotation of the axes.[†]

A similar situation exists for a four-dimensional space–time coordinate system. But, as well as the reversal of the direction of one, two, or three axes, we can here have a simultaneous reversal of all four axes (*four-dimensional inversion* or *four-inversion*). In purely mathematical terms, this transformation is a rotation of the four-coordinates. There is admittedly a specific difference between four-inversion and the rotations forming the Lorentz group, due to the fact that four-dimensional space–time geometry is pseudo-Euclidean. Because of this property, no physical (Lorentz) transformation of the frame of reference can bring the time axis outside the interior portions of the light cone (as defined in *Mechanics and Electrodynamics*, §34); physically, this expresses the impossibility of relative motion of two frames of reference at a speed exceeding that of light. Under four-inversion, however,

---

[†] Mathematically, the difference between the two types of linear transformation of the coordinates

$$x_i' = \sum_k \alpha_{ik} x_k$$

(where $x_1 = x$, $x_2 = y$, $x_3 = z$) is shown by the value of the determinant formed from their coefficients. For any rotation of the coordinates, the determinant $|\alpha_{ik}| = 1$, but for reflections that cannot be reduced to rotations $|\alpha_{ik}| = -1$.

the time axis (or rather each half of it) is brought from one interior portion of the light cone to the other.

Although this implies that four-inversion is physically impossible as a transformation of a reference frame, we can reasonably suppose that the difference from other four-dimensional rotations (Lorentz transformations) is unimportant in relation to the mathematical invariance of any particular equations. Thus we conclude that any relativistically invariant law of nature must also be invariant under four-inversion. It remains to ascertain what is the significance of this statement as regards the quantum theory of particle fields. This will be done for the simple case of a field of spin-zero particles.

In this case, the wave amplitudes $u(p)$ in the Ψ-operators (85.3) are scalars, and therefore independent of the sign of the argument, the four-momentum p^μ. Taking them outside the braces, we can therefore write simply

$$\Psi(t, \mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{p}} u\{\hat{a}_{\mathbf{p}} e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})} + \hat{b}_{\mathbf{p}}^{+} e^{i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})}\}. \tag{90.1}$$

Under four-inversion, t and $\mathbf{r}$ are replaced by $-t$ and $-\mathbf{r}$, and this expression becomes

$$\Psi(-t, -\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{p}} u\{\hat{a}_{\mathbf{p}} e^{i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})} + \hat{b}_{\mathbf{p}}^{+} e^{-i(\varepsilon t - \mathbf{p}\cdot\mathbf{r})}\}. \tag{90.2}$$

In the second quantisation formalism, the change from (90.1) to (90.2) must be expressed in terms of a transformation of the particle creation and annihilation operators. A comparison of (90.1) and (90.2) shows that this transformation consists of the interchange of the operators $\hat{a}_{\mathbf{p}}$ and $\hat{b}_{\mathbf{p}}^{+}$ or, equivalently, of the changes

$$\hat{a}_{\mathbf{p}} \to \hat{b}_{\mathbf{p}}^{+}, \quad \hat{b}_{\mathbf{p}} \to \hat{a}_{\mathbf{p}}^{+}. \tag{90.3}$$

The significance of the transformation (90.3) is evident. Inversion changes the sign of the momentum vector $\mathbf{p}$, but the sign of this is also changed by time reversal (the particle velocity changing to the opposite direction). Hence the simultaneous application of the transformations P and T leaves the particle momenta unaltered, and accordingly the operators transformed into each other pertain to states with the same $\mathbf{p}$. Next, time reversal, which replaces the future by the past, converts the appearance of a particle into its disappearance; accordingly, the particle creation and annihilation operators are interchanged. We also see, however, that in (90.3) the a and b operators are interchanged; this means that the transformation (90.3) includes the interchange of particles and antiparticles.

Thus, in the relativistic theory, there is a natural requirement of invariance under a transformation comprising spatial inversion, time reversal and charge conjugation. This is called the *CPT theorem.*[†]

By virtue of the *CPT* theorem, the violation of *CP* invariance in any phenomenon necessarily implies that of *T* invariance.

§91. **Neutrinos**

Dirac's equation is invariant with respect to inversion. This invariance occurs because the bispinor wave function includes two spinors which become each other on inversion. In turn, the need to include two spinors in the description of the particle arises from the mass of the particle: it is seen from (82.2) or (82.6) that the quantity m provides the "coupling" of these spinors in the wave equation.

The necessity disappears if the particle mass is zero. A particle of this kind with spin $\frac{1}{2}$ is the *neutrino.* The wave equation which describes such a particle can be derived from a single four-spinor, say the undotted spinor

$$\xi = \begin{pmatrix}\xi^1\\ \xi^2\end{pmatrix}.$$

The wave equation is

$$(\hat{p}_0 - \hat{\mathbf{p}}\cdot\boldsymbol{\sigma})\xi = 0, \tag{91.1}$$

i.e. the first equation (82.6) with $m = 0$.

For a plane wave (a particle with momentum $\mathbf{p}$ and energy ε), equation (91.1) reduces to the algebraic system

$$(\varepsilon - \mathbf{p}\cdot\boldsymbol{\sigma})\xi = 0.$$

[†] It was enunciated by G. Lüders, W. Pauli, and J. Schwinger (1955).

The energy of a zero-mass particle is related to its momentum by $\varepsilon = |\mathbf{p}|$. If $\mathbf{n}$ is a unit vector in the direction of motion, then

$$(\mathbf{n}\cdot\boldsymbol{\sigma})\xi = \xi. \tag{91.2}$$

This equation has a simple significance. For a two-component wave function, the matrix $\hat{\mathbf{s}} = \frac{1}{2}\boldsymbol{\sigma}$ is the particle spin operator (see §40). The product $\frac{1}{2}\mathbf{n}\cdot\boldsymbol{\sigma}$ is therefore the operator of the particle helicity λ (the component of the spin in the direction of the momentum). Hence equation (91.2) signifies that the particle has a definite helicity $\lambda = +\frac{1}{2}$, i.e. the spin is in the direction of motion.

Thus we conclude that a particle described by only one (undotted) spinor must always have a definite helicity $\lambda = +\frac{1}{2}$. In an exactly similar manner, for a particle described by the dotted spinor

$$\eta = \begin{pmatrix} \eta^{\dot{1}} \\ \eta^{\dot{2}} \end{pmatrix},$$

we have instead of (91.2) the equation

$$(\mathbf{n}\cdot\boldsymbol{\sigma})\eta = -\eta, \tag{91.3}$$

i.e. such a particle always has the helicity $\lambda = -\frac{1}{2}$, its spin being opposite to the momentum. We can say that in either case there must be longitudinal polarisation of the particle.

It is easily seen that a particle and antiparticle must have opposite helicities. If one of them is described by spinors ξ, the other must be described by the complex conjugate spinors ξ^*; this is evident from the form of the Ψ-operators (85.3), in which the particle and antiparticle annihilation operators $\hat{a}_\mathbf{p}$ and $\hat{b}_\mathbf{p}$ are multiplied by complex conjugate functions. But the spinor ξ^* conjugate to the undotted spinor ξ is equivalent to a dotted spinor, which proves the above statement. It is usual to call the particle with helicity $-\frac{1}{2}$ the neutrino, and that with helicity $+\frac{1}{2}$ the antineutrino.[†]

[†] The existence of neutrinos as electrically neutral particles with mass zero and spin $\frac{1}{2}$ was theoretically predicted by W. Pauli (1931) in order to explain the properties of β-decay. The theory of the neutrino as a particle described by a four-spinor with two components was evolved by L. D. Landau, A. Salam, T. D. Lee, and C. N. Yang (1957).

Inversion changes the sign of the helicity since the projection of the spin on the direction of motion is obtained by scalar multiplication of the angular momentum vector by the particle momentum. The angular momentum vector, being an axial vector, is unchanged by inversion; the momentum, a polar vector, changes sign. This clearly shows that the neutrino is not symmetrical with respect to inversion: inversion "converts" the neutrino into a particle that does not exist in nature: a neutrino with reversed helicity. There is symmetry only with respect to combined inversion, i.e. inversion with simultaneous replacement of the neutrino by an antineutrino. It is therefore natural that mirror symmetry does not occur in processes involving neutrinos, such as the β-decay of the neutron into a proton, an electron and an antineutrino ($n \rightarrow p+e+\bar{\nu}$).

CHAPTER 14

ELECTRONS IN AN EXTERNAL FIELD

§92. Dirac's equation for an electron in an external field

The wave equations of free particles express essentially only those properties which depend on the general requirements of space–time symmetry. Physical processes involving the particles, however, depend on their interaction properties.

In the relativistic theory it proves impossible to obtain by any simple generalisation of the wave equations a description of particles that are capable of strong interactions, i.e. a description going beyond the information contained in the equations for free particles.

The wave-equation method, however, is applicable to the description of electromagnetic interactions of particles that are not capable of strong interactions. These include electrons (and positrons), and the very wide domain of electron quantum electrodynamics is therefore accessible to the existing theory.†

In this chapter we shall discuss problems of quantum electrodynamics which fall within the scope of single-particle theory. These are problems in which the number of particles is unchanged, and the interaction can be represented in terms of an external electromagnetic field due to sources whose state remains constant in the course of time.

---

† There are also unstable particles, the muons, which are not capable of strong interactions; they have the same spin ($\frac{1}{2}$) as the electron, and are described by the same quantum electrodynamics as regards phenomena occurring in times short in comparison with their lifetime (with respect to weak interactions).

The wave equation for an electron in a given external field can be derived in the same way as in the non-relativistic theory (§43). Let Φ be the scalar potential and **A** the vector potential of the field. We obtain the desired equation on replacing the momentum operator $\hat{\mathbf{p}} = -i\nabla$ in the Hamiltonian of Dirac's equation (83.9) by $\hat{\mathbf{p}}-e\mathbf{A}$ and adding to the Hamiltonian the potential energy $e\Phi$ of the particle:[†]

$$\hat{H} = \boldsymbol{\alpha}\cdot(\hat{\mathbf{p}}-e\mathbf{A})+\beta m+e\Phi. \tag{92.1}$$

These are the only necessary changes; no artificially added terms like that in (43.4) are needed here. We shall see later that the magnetic moment of the electron makes its appearance automatically.

In four-dimensional notation, the transition from (83.9) to (92.1) signifies the replacement of the four-momentum operator $\hat{p}_\mu = i\,\partial/\partial x^\mu$ according to

$$\hat{p}_\mu \to \hat{p}_\mu - eA_\mu, \tag{92.2}$$

where $A^\mu = (\Phi, \mathbf{A})$, $A_\mu = (\Phi, -\mathbf{A})$ is the four-potential of the field. Hence Dirac's equation for a particle in a field can also be written

$$[\gamma^\mu(\hat{p}_\mu - eA_\mu)-m]\Psi = 0, \tag{92.3}$$

which is obtained from (83.3) by making the same change.

The current density expressed in terms of the wave function is given by the same formula (84.7) as when the external field is absent. It is easily seen that, on repeating with (92.3) the same arguments as were used in the derivation of (84.7), the four-potential A_μ disappears and the equation of continuity is obtained for the previous expression for the current.

§93. Magnetic moment of the electron[‡]

In §43 we have established the form of the non-relativistic Hamiltonian for the motion of a particle with spin in an external magnetic field. In that expression, however, the magnetic moment of the particle

---

[†] The letter e denotes the charge together with its sign, so that for the electron $e = -|e|$, but for the positron $e = +|e|$.

[‡] In §§93 and 94 ordinary units are used.

appeared as an empirical parameter whose value could not be found theoretically. For a particle whose behaviour in an electromagnetic field obeys Dirac's equation (92.3), such as an electron, the value of the magnetic moment is established by the equation itself.

For this purpose, we shall show how Dirac's equation may be put in an approximate form corresponding to the non-relativistic Hamiltonian (43.4). Since the particle is regarded as moving at velocities $v \ll c$, we can reasonably start from the standard representation of the bispinor function Ψ, in which one pair of components is small in comparison with the other, $\chi \ll \phi$ (see the end of §83).

In §83, Dirac's equations in the standard representation of the wave function were written for a free particle (83.11). The introduction of an external electromagnetic field into these equations is achieved by a change of operators in accordance with (92.2); thus we have

$$\left.\begin{aligned}(\hat{p}_0-e\Phi)\phi-\boldsymbol{\sigma}.(\hat{\mathbf{p}}-e\mathbf{A}/c)\chi &= mc\phi,\\ -(\hat{p}_0-e\Phi)\chi+\boldsymbol{\sigma}.(\hat{\mathbf{p}}-e\mathbf{A}/c)\phi &= mc\chi,\end{aligned}\right\} \tag{93.1}$$

where

$$\hat{p}_0 = (i\hbar/c)\,\partial/\partial t, \quad \hat{\mathbf{p}} = -i\hbar\nabla.$$

To obtain the non-relativistic approximation, however, the wave function must first be subjected to a further transformation. The reason is that the relativistic expression for the particle energy (and therefore the relativistic Hamiltonian) contains an extra term, the rest energy mc^2, in comparison with the non-relativistic expression. This gives rise to an extra factor $\exp(-imc^2t/\hbar)$ in the time dependence of the wave function. To exclude this factor, we replace Ψ by a new wave function Ψ':

$$\Psi = \Psi' \exp(-imc^2t/\hbar). \tag{93.2}$$

Substituting this in (93.1), we obtain the following equations for the two-component quantities ϕ' and χ' which constitute the four-component Ψ':

$$\left(i\hbar\frac{\partial}{\partial t}-e\Phi\right)\phi' = c\boldsymbol{\sigma}.(\hat{\mathbf{p}}-e\mathbf{A}/c)\chi', \tag{93.3}$$

$$\left(i\hbar\frac{\partial}{\partial t}-e\Phi+2mc^2\right)\chi' = c\boldsymbol{\sigma}.(\hat{\mathbf{p}}-e\mathbf{A}/c)\phi'. \tag{93.4}$$

In what follows we shall omit the primes from ϕ' and χ'; this will cause no misunderstanding, since only the transformed function Ψ' is used in the present section.

In the first approximation, only the largest term $2mc^2$ is retained in the parenthesis on the left of (93.4). Then this equation enables us to express χ directly in terms of ϕ:

$$\chi = \frac{1}{2mc}\,\boldsymbol{\sigma}\,.(\hat{\mathbf{p}}-e\mathbf{A}/c)\phi. \tag{93.5}$$

The factor $1/c$ on the right expresses the smallness of χ in comparison with ϕ. Now, substituting (93.5) in (93.3), we obtain an equation containing only ϕ:

$$\left(i\hbar\frac{\partial}{\partial t}-e\Phi\right)\phi = \frac{1}{2m}[\boldsymbol{\sigma}\,.(\hat{\mathbf{p}}-e\mathbf{A}/c)]^2\,\phi.$$

In expanding the right-hand side, we use the following properties of the Pauli matrices, which follow immediately from the definitions (82.5):

$$\left.\begin{aligned} \sigma_x^2 &= \sigma_y^2 = \sigma_z^2 = 1,\\ \sigma_y\sigma_z &= -\sigma_z\sigma_y = i\sigma_x,\\ \sigma_z\sigma_x &= -\sigma_x\sigma_z = i\sigma_y,\\ \sigma_x\sigma_y &= -\sigma_y\sigma_x = i\sigma_z. \end{aligned}\right\} \tag{93.6}$$

Writing temporarily $\hat{\mathbf{f}} = \hat{\mathbf{p}}-e\mathbf{A}/c$, we have

$$\begin{aligned}(\boldsymbol{\sigma}\,.\hat{\mathbf{f}})^2 &= (\sigma_x\hat{f}_x+\sigma_y\hat{f}_y+\sigma_z\hat{f}_z)(\sigma_x\hat{f}_x+\sigma_y\hat{f}_y+\sigma_z\hat{f}_z)\\ &= \hat{f}_x^2+\hat{f}_y^2+\hat{f}_z^2+i\sigma_z(\hat{f}_x\hat{f}_y-\hat{f}_y\hat{f}_x)+\ \dots\ .\end{aligned}$$

If $\hat{f}_x$, $\hat{f}_y$, and $\hat{f}_z$ commuted, this would be simply $\hat{\mathbf{f}}^2$, but in the present case

$$\begin{aligned}\hat{f}_x\hat{f}_y-\hat{f}_y\hat{f}_x &= \left(-i\hbar\frac{\partial}{\partial x}-eA_x/c\right)\left(-i\hbar\frac{\partial}{\partial y}-eA_y/c\right)\\ &\quad-\left(-i\hbar\frac{\partial}{\partial y}-eA_y/c\right)\left(-i\hbar\frac{\partial}{\partial x}-eA_x/c\right)\\ &= \frac{ie\hbar}{c}\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)\\ &= (ie\hbar/c)H_z, \quad \text{etc.,}\end{aligned}$$

where $\mathbf{H} = \operatorname{curl} \mathbf{A}$ is the magnetic field. Thus

$$[\boldsymbol{\sigma}\cdot(\hat{\mathbf{p}}-e\mathbf{A}/c)]^2 = (\hat{\mathbf{p}}-e\mathbf{A}/c)^2-(e\hbar/c)\boldsymbol{\sigma}\cdot\mathbf{H},$$

and we therefore obtain the following equation for the two-component wave function ϕ:

$$i\hbar\frac{\partial\phi}{\partial t} = \left[\frac{1}{2m}(\hat{\mathbf{p}}-e\mathbf{A}/c)^2-\frac{e\hbar}{2mc}\boldsymbol{\sigma}\cdot\mathbf{H}+e\Phi\right]\phi \equiv \hat{H}\phi. \quad (93.7)$$

This is *Pauli's equation.* A comparison of the Hamiltonian in it with (43.4) shows that the electron has a magnetic moment corresponding to the operator

$$\hat{\mu} = \frac{e\hbar}{2mc}\boldsymbol{\sigma} = \frac{e\hbar}{mc}\hat{\mathbf{s}}, \quad (93.8)$$

where $\hat{\mathbf{s}} = \frac{1}{2}\boldsymbol{\sigma}$ is the electron spin operator. The magnitude of this angular momentum is, by (43.1),

$$\mu = e\hbar/2mc. \quad (93.9)$$

As already mentioned in §43, the gyromagnetic ratio e/mc for the intrinsic magnetic moment of the electron is twice its value for a magnetic moment due to orbital motion.[†]

Formula (93.9) is valid also for the magnetic moment of the muon (with the latter's mass as m in the denominator). It is, however, entirely incorrect for protons and neutrons, although these are also spin-$\frac{1}{2}$ particles. The difference is particularly noticeable for the neutron, which, being electrically neutral, should according to (93.9) have no magnetic moment. This shows clearly the inapplicability of the existing quantum electrodynamics to particles that are capable of strong interactions.

[†] This result was derived by P. A. M. Dirac in 1928. The two-component wave function satisfying (93.7) was introduced by W. Pauli (1927), before Dirac's discovery of his equation.

§94. Spin–orbit interaction

The calculations given in §93 represent essentially the beginning of an expansion of the exact solution of Dirac's equation in powers of the small ratio v/c. Equation (93.7) corresponds to including in such an expansion only terms of the first order of smallness, as is indicated by the factor $1/c$ in the additional term $-\hat{\boldsymbol{\mu}}.\mathbf{H}$ in the Hamiltonian.

In the second approximation, the Hamiltonian contains further terms. The corresponding calculations are more laborious, however, and will not be given here; we shall simply state the final result for the Hamiltonian of an electron in an external electric field, as far as terms of order $1/c^2$:

$$\hat{H} = \frac{\hat{\mathbf{p}}^2}{2m}+e\Phi-\frac{\hat{\mathbf{p}}^4}{8m^3c^2}-\frac{e\hbar}{4m^2c^2}\boldsymbol{\sigma}.\mathbf{E}\times\hat{\mathbf{p}}-\frac{e\hbar^2}{8m^2c^2}\operatorname{div}\mathbf{E}, \qquad (94.1)$$

where Φ is the potential and $\mathbf{E} = -\operatorname{grad}\Phi$ is the field strength. As in (93.7), this Hamiltonian acts on a two-component wave function.

The last three terms in (94.1) are the required corrections of order $1/c^2$. The first of them corresponds to the relativistic correction to the classical expression for the kinetic energy of the particle:

$$\sqrt{(c^2\mathbf{p}^2+m^2c^4)}-mc^2 \approx \frac{\mathbf{p}^2}{2m}-\frac{\mathbf{p}^4}{8m^3c^2}+\dots .$$

The next correction term in (94.1) may be called the *spin–orbit interaction energy*: it is the energy of the interaction of the moving magnetic moment with the electric field. If the electric field is centrally symmetric, then

$$\mathbf{E} = -\frac{\mathbf{r}}{r}\frac{\mathrm{d}\Phi}{\mathrm{d}r},$$

and the spin–orbit interaction operator can be put in the form

$$\hat{V}_{sl} = \frac{e\hbar}{4m^2c^2r}\boldsymbol{\sigma}.\mathbf{r}\times\hat{\ }\frac{\mathrm{d}\Phi}{\mathrm{d}r} = \frac{\hbar^2}{2m^2c^2r}\frac{\mathrm{d}U}{\mathrm{d}r}\hat{\mathbf{l}}.\hat{\mathbf{s}}. \qquad (94.2)$$

Here $\hbar\hat{\mathbf{l}} = \mathbf{r}\times\hat{\mathbf{p}}$ is the electron orbital angular momentum operator,

$\mathbf{s} = \frac{1}{2}\boldsymbol{\sigma}$ is the electron spin operator, and $U = e\phi$ is the potential energy of the electron in the field. An interaction of this type has already been considered in §51 as one source of the fine structure of atomic energy levels.[†]

The last correction term in (94.1) is zero except at points where there are charges creating the field, since div $\mathbf{E} = 0$ except at such points.

The Hamiltonian (94.1) can be used to calculate the relativistic corrections to the energy levels of the hydrogen atom, i.e. of an electron in the Coulomb field of a fixed proton nucleus with charge $+|e|$.

The field potential of the charge $+|e|$ is $\Phi = |e|/r$, and the divergence of the field is div $\mathbf{E} = -\triangle\Phi = 4\pi|e|\,\delta(\mathbf{r})$; cf. *Mechanics and Electrodynamics*, (59.10). The correction terms in the Hamiltonian of the hydrogen atom, which we denote jointly by $\hat{V}^{(2)}$, have the form

$$\hat{V}^{(2)} = \frac{\hbar^2}{8m^3c^2}\triangle^2 + \frac{\hbar^2e^2}{2m^2c^2r^3}\,\hat{\mathbf{l}}\,.\,\hat{\mathbf{s}} + \frac{\pi e^2\hbar^2}{2m^2c^2}\,\delta(\mathbf{r}). \tag{94.3}$$

The non-relativistic expression for the energy levels of the hydrogen atom is (§31)

$$E_{\text{non-r}} = -me^4/2\hbar^2n^2; \tag{94.4}$$

it depends only on the principal quantum number n, and not on the orbital angular momentum l of the electron, which, for a given n, takes the values $l = 0, 1, \ldots, n-1$. The non-relativistic levels (94.4) are also independent of the direction of the electron spin relative to its orbital angular momentum, i.e. independent of the total angular momentum j, which, for a given $l \neq 0$, can take two values: $j = l \pm \frac{1}{2}$.

The required corrections ΔE to the levels (94.4) can be found from the general rules of perturbation theory (§32). Regarding (94.3) as the operator of a small perturbation, we have to calculate its mean value (diagonal matrix element) with respect to the unperturbed wave functions, i.e. the ordinary non-relativistic wave functions of the

---

[†] Another type of relativistic interaction, the spin–spin interaction, of course occurs only in a system of more than one particle, and not for a single electron in an external field.

hydrogen atom. Calculation gives the result

$$\Delta E = -\left(\frac{1}{j+\frac{1}{2}} - \frac{3}{4n}\right)\frac{me^4\alpha^2}{2\hbar^2 n^3}, \tag{94.5}$$

where

$$\alpha = e^2/\hbar c = 1/137.04 \tag{94.6}$$

and is called the *fine-structure constant*.[†] The factor α^2 expresses the smallness of the correction (94.5) relative to (94.4).

The shift (94.5) of the level depends on j as well as on n. This dependence corresponds to the splitting of the levels (94.4) into fine-structure components; the degeneracy present in the non-relativistic approximation is said to have been removed. This removal is not complete, however, since the levels with the same values of n and j but different values of $l = j \pm \frac{1}{2}$ remain doubly degenerate. (This again is a property peculiar to the hydrogen atom with the purely Coulomb field of its nucleus, and does not occur for more complex atoms.) Thus the sequence of hydrogen levels, with allowance for the fine structure, is

$$\begin{aligned}&1s_{1/2};\\ &\underbrace{2s_{1/2},\ 2p_{1/2}},\ 2p_{3/2};\\ &\underbrace{3s_{1/2},\ 3p_{1/2}},\ \underbrace{3p_{3/2},\ 3d_{3/2}},\ 3d_{5/2};\\ &\ldots,\end{aligned}$$

where the braces join the mutually degenerate states. Only the levels having the largest possible value of j (for a given m) remain non-degenerate.

To anticipate, it may be mentioned here that the remaining degeneracy is removed by "radiative corrections" (the *Lamb shift*), which are neglected in Dirac's equations for the single-electron problem. These corrections will be discussed in §106.

[†] This formula was first derived by A. Sommerfeld on the basis of the old Bohr theory, before the discovery of quantum mechanics.

CHAPTER 15

RADIATION

§95. The electromagnetic interaction operator

Let us now proceed from problems in which the electromagnetic field has the passive role of providing external conditions for the motion of particles to the broader category of electrodynamic phenomena accompanied by a change in the state of the field itself. These are phenomena involving the emission, absorption, or scattering of photons by systems of charged particles.

The interaction of electrons with a field of electromagnetic radiation can, as a rule, be treated by means of perturbation theory. This is because the electromagnetic interactions are comparatively weak. The interaction of an electron with the field is determined by its charge e, and the "coupling constant" that gives the scale of the interaction is the dimensionless quantity $\alpha = e^2/\hbar c$ already mentioned in §94 as the fine-structure constant. The weakness of the electromagnetic interactions is expressed by the smallness of this constant: $\alpha = 1/137$, a smallness that is of fundamental importance in quantum electrodynamics.

Let us first ascertain the form of the operator for the interaction of an electron with the radiation field, which acts as the perturbation operator. We shall suppose (as in Chapter 11) that the field potentials are taken in a gauge such that the scalar potential $\Phi = 0$, so that the field is described by just the vector potential $\mathbf{A}$. According to (92.1), the interaction of an electron with a given electromagnetic field is described by a term $\hat{V} = -e\boldsymbol{\alpha}.\mathbf{A}$ in its Hamiltonian. To go to the more general case of processes in which the state of the field changes, the potential **A** must be replaced by the corresponding second-quantised operator $\hat{\mathbf{A}}$; then the interaction operator is[†]

$$\hat{V} = -e\boldsymbol{\alpha}\cdot\hat{\mathbf{A}}. \tag{95.1}$$

The operator $\hat{\mathbf{A}}$ is the sum

$$\hat{\mathbf{A}}(t, \mathbf{r}) = \sum_n \left\{\hat{c}_n \mathbf{A}_n(t, \mathbf{r}) + \hat{c}_n^+ \mathbf{A}_n^*(t, \mathbf{r})\right\}, \tag{95.2}$$

which contains the operators of photon annihilation and creation in various states labelled by the suffix n; the coefficients $\mathbf{A}_n(t, \mathbf{r})$ act as the wave functions of these states. The state of the field is specified by the set of occupation numbers N_n of all the photon states. The photon states themselves may be specified in various ways, depending on the particular problem concerned. For example, if we are considering the emission or absorption of photons with definite wave vectors **k** and polarisations **e**, the wave functions $\mathbf{A}_n(t, \mathbf{r})$ are the plane waves (76.16). If, on the other hand, the problem concerns the emission of photons with definite values of the angular momentum j, the $\mathbf{A}_n$ are spherical waves, as discussed in §78.

In the first approximation of perturbation theory, the probability of a particular process is given by $|V_{fi}|^2$, where V_{fi} is the matrix element of the perturbation operator for a transition between the initial (i) and final (f) states of the system of charges and the field. Each operator $\hat{c}_n$, $\hat{c}_n^+$ has non-zero matrix elements only for an increase or decrease of the corresponding occupation number N_n by 1 (the other occupation numbers remaining unchanged). The operator $\hat{\mathbf{A}}$ therefore also has matrix elements only for transitions in which the number of photons changes by 1. That is, only processes of the emission or absorption of a single photon occur in the first approximation of perturbation theory.

[†] The charge-conjugation operator, which replaces particles by antiparticles, must not affect the form of the interaction operator. It replaces positively charged particles by negatively charged ones, and in particular makes the change $e \rightarrow -e$. The invariance of $\hat{V}$ requires a simultaneous change of the photon field operator $\hat{\mathbf{A}} \rightarrow -\hat{\mathbf{A}}$; photons are therefore charge-odd particles.

According to (76.12), the matrix elements are

$$\langle N_n - 1 | c_n | N_n \rangle = \sqrt{N_n}, \tag{95.3}$$

$$\langle N_n + 1 | c_n^+ | N_n \rangle = \sqrt{(N_n + 1)}. \tag{95.4}$$

The first of these corresponds to the absorption of one photon (of type n), the occupation number decreasing by one; the second corresponds to the emission of one photon, the occupation number increasing by one. If there are no photons of type n in the initial state of the field, then $\langle 1 | c_n^+ | 0 \rangle = 1$; the matrix element of the operator $\hat{\mathbf{A}}$ also contains the factor $\mathbf{A}_n^*$ that appears in the sum (95.2) as the coefficient of $\hat{c}_n^+$. Thus the complete matrix element of the operator (95.1) for the emission of a photon is

$$V_{fi}(t) = -e \int (\Psi_f^* \boldsymbol{\alpha} \Psi_i) \cdot \mathbf{A}_n^* \, \mathrm{d}V, \tag{95.5}$$

where Ψ_i and Ψ_f are the wave functions of the initial and final states of the emitting electron.[†] Similarly, we obtain the matrix element for photon absorption:

$$V_{fi}(t) = -e \int (\Psi_f^* \boldsymbol{\alpha} \Psi_i) \cdot \mathbf{A}_n \, \mathrm{d}V. \tag{95.6}$$

This differs from (95.5) only by having $\mathbf{A}_n$ in place of $\mathbf{A}_n^*$.

The argument t of V_{fi} is shown in order to emphasise that the matrix element is time-dependent. By separating the time factor in the wave functions, we can change in the usual way, in accordance with the rule (11.4), to the time-independent matrix elements:

$$V_{fi}(t) = V_{fi} \exp[-i(E_i - E_f \mp \omega)t], \tag{95.7}$$

where E_i, E_f are the initial and final energies of the radiating system, and the upper and lower signs in the exponent are for emission and absorption of a photon with energy ω.

[†] To avoid misunderstanding, it should be stressed that one electron can radiate only when moving in an external field. The impossibility of photon emission by a free electron (moving with constant velocity) is especially clear if the electron is considered in a frame of reference where it is at rest; in that frame, the energy of the electron is m, and cannot decrease as it would have to do on the emission of a photon.

The product

$$\mathbf{j}_{fi} = \Psi_f^* \boldsymbol{\alpha} \Psi_i \tag{95.8}$$

which appears in the integrand in (95.5) and (95.6) is similar in form to the expression $\mathbf{j} = \Psi^* \boldsymbol{\alpha} \Psi$ (84.9) for the current in Dirac's equation; it has two different (initial and final) wave functions instead of the same one twice. The quantity (95.8) is called the *transition current.*

For the emission (or absorption) of a photon having a definite direction of the wave vector **k** and a definite polarisation **e**, the function $\mathbf{A}_n(\mathbf{r})$ must be taken as

$$\mathbf{A}_n(\mathbf{r}) = \mathbf{e}\sqrt{\frac{2\pi}{\omega\Omega}}\, e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{95.9}$$

i.e. the plane wave (76.16) without the factor $e^{-i\omega t}$. For the matrix element of a transition with the emission of such a photon, we have

$$V_{fi} = -e\sqrt{\frac{2\pi}{\omega\Omega}}\, \mathbf{e}^* \cdot \mathbf{j}_{fi}(\mathbf{k}), \tag{95.10}$$

where

$$\mathbf{j}_{fi}(\mathbf{k}) = \int \mathbf{j}_{fi}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}}\, \mathrm{d}V. \tag{95.11}$$

The integral (95.11) is the Fourier component of the function $\mathbf{j}_{fi}(\mathbf{r})$, and is called the *transition current in the momentum representation.*

The photon emission probability can be found directly from the matrix element (95.10) by means of a general formula of perturbation theory derived in §35. We shall suppose that the initial and final states of the emitter belong to its discrete spectrum of energy levels. The final state of the electron + field system will, however, belong to the continuous spectrum, because of the emitted photon, the spectrum of possible values of the photon energy being continuous. Thus we have here the same formulation of the problem as was considered in §35. According to (35.6), the probability (per unit time) of the transition $i \to f$ with emission of a photon is

$$\mathrm{d}w = 2\pi |V_{fi}|^2 \delta(E_i - E_f - \omega)\, \mathrm{d}\nu, \tag{95.12}$$

where ν arbitrarily denotes the ensemble of quantities describing the state of the photon and taking a continuous sequence of values. For photons with definite values of the wave vector, the quantities ν are the components of **k**, so that $d\nu = dk_x \, dk_y \, dk_z = \omega^2 \, d\omega \, do$ (where do is the element of solid angle for the directions of **k**). With this choice of ν in formula (95.12), it would be assumed that the photon wave function is normalised to $\delta(\mathbf{k})$. But the function (95.9) is normalised to one photon in the volume Ω; with this normalisation, the wave function contains a factor $1/\sqrt{\Omega}$ instead of the factor $(2\pi)^{-3/2}$ in the normalisation to $\delta(\mathbf{k})$ (cf. (27.9) and (12.10)). Hence formula (95.12) must now be written

$$dw = 2\pi \, |V_{fi}|^2 \, \delta(E_i - E_f - \omega) \frac{\Omega \omega^2 \, d\omega \, do}{(2\pi)^3}. \tag{95.13}$$

The δ-function here expresses the law of conservation of energy: the energy of the emitted photon is equal to the decrease in the energy of the emitter, $\omega = E_i - E_f$. The integration of (95.13) with respect to ω removes this δ-function, and gives the following final expression for the probability of emission of a photon with energy $\omega = E_i - E_f$ in a direction in the solid angle do:

$$dw = \frac{\Omega}{4\pi^2} \, |V_{fi}|^2 \, \omega^2 \, do. \tag{95.14}$$

In this we must substitute the matrix element (95.10).

§96. Spontaneous and stimulated emission[†]

In subsequent sections we shall use the formulae derived above to calculate the transition probabilities in various specific cases. Here we shall consider certain general relations between radiative processes of various kinds.

The matrix element (95.5) pertains to the emission of a photon with the condition that there is no photon of the same type in the initial state of the field. If there are already N_n such photons in the initial

---

[†] In this section, ordinary units are used.

state, the transition matrix element is multiplied by $\sqrt{(N_n+1)}$, according to (95.4). The transition probability is accordingly multiplied by N_n+1. The 1 in this factor corresponds to the spontaneous emission, which occurs even if $N_n = 0$. The term N_n gives rise to the *stimulated* or *induced* emission: we see that the presence of photons in the initial state of the field stimulates the further emission of photons of the same kind.

If the transition $i \to f$ represents the emission of a photon by the system as it goes from a level E_i to a lower level E_f, the reverse transition $f \to i$ will represent the absorption of a similar photon by the system as it goes from the level E_f to the level E_i. The matrix element of this reverse transition differs from that of the original transition in that the factor (95.4) is replaced by (95.3), i.e. $\sqrt{(N_n+1)}$ is replaced by $\sqrt{N_n}$. Hence it follows that the photon emission and absorption probabilities (for transitions between a given pair of levels of the radiating system) are related by

$$w_e/w_a = (N_n+1)/N_n, \tag{96.1}$$

an expression first derived in 1916 by A. Einstein, who thus predicted the phenomenon of stimulated emission.

The number of photons can be related to the intensity of the external radiation incident on the system. Let

$$I_{\mathbf{k}\mathbf{e}}\, \mathrm{d}\omega\, \mathrm{d}o \tag{96.2}$$

be the radiation energy incident on unit area per unit time and having polarisation $\mathbf{e}$, frequency in the range $\mathrm{d}\omega$, and wave-vector direction in the element of solid angle $\mathrm{d}o$. These ranges correspond to $\Omega k^2\, \mathrm{d}k\, \mathrm{d}o/(2\pi)^3$ field oscillators in the volume Ω, each having $N_{\mathbf{k}\mathbf{e}}$ photons of the specified polarisation. Hence the same energy (96.2) is given by the product

$$\frac{c}{\Omega}\,\frac{\Omega k^2\, \mathrm{d}k\, \mathrm{d}o}{(2\pi)^3} N_{\mathbf{k}\mathbf{e}}\hbar\omega = \frac{\hbar\omega^3}{8\pi^3c^2} N_{\mathbf{k}\mathbf{e}}\, \mathrm{d}\omega\, \mathrm{d}o.$$

Hence we find the required relation

$$N_{\mathbf{k}\mathbf{e}} = \frac{8\pi^3c^2}{\hbar\omega^3} I_{\mathbf{k}\mathbf{e}}\,. \tag{96.3}$$

Let $dw_{\mathbf{k}e}^{(sp)}$ be the probability of spontaneous emission of a photon with polarisation **e** into the solid angle do, and let the indices (in) and (a) denote the corresponding probabilities for induced emission and for absorption. According to (96.1) and (96.3) these probabilities are related as follows:

$$dw_{\mathbf{k}e}^{(a)} = dw_{\mathbf{k}e}^{(in)} = dw_{\mathbf{k}e}^{(sp)} \cdot \frac{8\pi^3 c^2}{\hbar\omega^3} I_{\mathbf{k}e} . \tag{96.4}$$

If the incident radiation is isotropic and unpolarised ($I_{\mathbf{k}e}$ independent of the directions of **k** and **e**), then the integration of (96.4) with respect to do and summation with respect to the polarisation gives similar relations between the total probabilities of radiative transitions (between given states i and f of the system of charges):

$$w^{(a)} = w^{(in)} = w^{(sp)} \cdot \frac{\pi^2 c^2}{\hbar\omega^3} I, \tag{96.5}$$

where $I = 2 \times 4\pi I_{\mathbf{k}e}$ is the total spectral intensity of the incident radiation.

If the states i and f of the emitting (or absorbing) system are degenerate,[†] the total probability of emission (or absorption) of the photons concerned is found by summation over all mutually degenerate final states and averaging over all possible initial states. Let the degrees of degeneracy (the *statistical weights*) of the states i and f be g_i and g_f. For processes of spontaneous or induced emission, the states i are the initial states, and for absorption the states f. Assuming in each case that all g_i or g_f initial states are equally probable, we obviously have instead of (96.5) the relations

$$g_f w^{(a)} = g_i w^{(in)} = g_i w^{(sp)} \frac{\pi^2 c^2}{\hbar\omega^3} I. \tag{96.6}$$

[†] This may be, for example, degeneracy with respect to the directions of the angular momentum of the radiating atom.

§97. Dipole radiation

A very important case is that where the photon wavelength λ is large compared with the dimensions a of the radiating system. This situation is generally caused by the particles' having a velocity small compared with that of light (cf. *Mechanics and Electrodynamics*, §82).

In the first approximation with respect to the small ratio a/λ, we can replace the factor $e^{-i\mathbf{k}\cdot\mathbf{r}}$ by unity in the integral (95.11), since it varies only slightly over the dimensions of the system, i.e. in the region where the functions ψ_i and ψ_f are appreciably different from zero. This change thus implies that the photon momentum is neglected in comparison with the momenta of the particles in the system (the photon momentum being $\hbar\mathbf{k}$ in ordinary units, and the particle momenta being of the order of $\hbar/a$). This approximation corresponds to the dipole case in classical radiation theory.

In the same approximation, the integral

$$\mathbf{j}_{fi}(0) = \int \psi_f^* \boldsymbol{\alpha}\psi_i \, dV$$

may be replaced by the non-relativistic expression, i.e. simply by the matrix element of the electron velocity $\mathbf{v}$ with respect to the Schrödinger (non-relativistic) wave functions. In turn, the matrix element $\mathbf{v}_{fi}$ can be expressed in terms of the corresponding matrix element for the electron radius vector: since $\mathbf{v} = \dot{\mathbf{r}}$, according to (11.8) we have $\mathbf{v}_{fi} = i(E_f - E_i)\mathbf{r}_{fi}$. The difference $E_f - E_i$ is equal to the frequency ω of the photon emitted, so that

$$\mathbf{j}_{fi} = \mathbf{v}_{fi} = -i\omega\mathbf{r}_{fi} = -(i\omega/e)\mathbf{d}_{fi}, \tag{97.1}$$

where $\mathbf{d} = e\mathbf{r}$ is the dipole moment of the electron (in its orbital motion). Substituting (97.1) in (95.10), we find[†]

$$V_{fi} = i\sqrt{\frac{2\pi\omega}{\Omega}}\,\mathbf{e}^* \cdot \mathbf{d}_{fi}, \tag{97.2}$$

[†] A similar expression exists for the matrix element of the transition with photon absorption:

$$V_{fi} = -i\sqrt{\frac{2\pi\omega}{\Omega}}\,\mathbf{e} \cdot \mathbf{d}_{fi}. \tag{97.2a}$$

This is obtained from (95.6) in exactly the same way as (97.2) from (95.5).

and then, using (95.14), we have the following formula for the dipole emission probability:

$$dw = \frac{\omega^3}{2\pi} |\mathbf{e}^* . \mathbf{d}_{fi}|^2 \, do. \tag{97.3}$$

The direction of the photon wave vector **k** appears implicitly here, since the polarisation vector **e** must be perpendicular to **k**.

The total emission probability is obtained by integrating (97.3) over all directions of the photon and summing over its two possible independent polarisations. Let **e** correspond to linear polarisation; then **e** is a real unit vector, and the product $\mathbf{e}^* . \mathbf{d}_{fi}$ is a Cartesian component of the vector $\mathbf{d}_{fi}$. Replacing the square $|(\mathbf{d}_{fi})_x|^2$ by its mean value $\frac{1}{3}|\mathbf{d}_{fi}|^2$, we reduce the subsequent integration over do to a simple multiplication by 4π, and the summation over polarisations to a multiplication by 2. Thus the total probability of photon emission is

$$w = (4\omega^3/3)\, |\mathbf{d}_{fi}|^2,$$

or, in ordinary units,

$$w = (4\omega^3/3\hbar c^3)\, |\mathbf{d}_{fi}|^2. \tag{97.4}$$

The intensity I is found by multiplying the probability by $\hbar\omega$:

$$I = (4\omega^4/3c^3)\, |\mathbf{d}_{fi}|^2. \tag{97.5}$$

It should be noted that the approximate expression (97.2) for the matrix element is the matrix element of the operator

$$\hat{V} = -\hat{\mathbf{E}} . \mathbf{d}, \tag{97.6}$$

where $\hat{\mathbf{E}} = -\partial\hat{\mathbf{A}}/\partial t$ is the electric field operator and **d** the electron dipole moment operator; (97.2) is obtained from (97.6) in exactly the same way as (95.5) from (95.1). The approximate interaction operator (97.6) corresponds precisely to the classical non-relativistic expression for the potential energy of a system of charges in a quasi-uniform electric field (see *Mechanics and Electrodynamics*, §64). This is an important point, in that it allows a wide range of applications of the formulae derived in this section; it applies not only to a single-electron radiator but also to radiation from any non-relativistic system of particles.

Formula (97.5) shows a direct analogy to the classical formula

(*Mechanics and Electrodynamics*, (82.12)) for the intensity of dipole radiation from a system of particles in periodic motion: the intensity of radiation with frequency $\omega = n\omega_0$ (where ω_0 is the frequency of the particle motion and n is an integer) is

$$I_n = (4\omega^4/3c^3)\,|\mathbf{d}_n|^2, \tag{97.7}$$

where the $\mathbf{d}_n$ are the components in the Fourier expansion of the dipole moment of the system:

$$\mathbf{d}(t) = \sum_{n=-\infty}^{\infty} \mathbf{d}_n e^{-in\omega_0 t}. \tag{97.8}$$

The quantum formula (97.5) is obtained from (97.7) by replacing these Fourier components by the matrix elements of the corresponding transitions. This rule (which is an expression of Bohr's *correspondence principle*) is a particular case of a general relation between the Fourier components of classical quantities and the quantum matrix elements in the quasi-classical case (§27). The radiation is quasi-classical for transitions between states having large quantum numbers; the photon energy $\hbar\omega = E_i - E_f$ is then small in comparison with the energies E_i and E_f of the radiator. But the exact formula (97.5) (which does not depend on the assumption of the quasi-classical case) has the same form for any value of ω, small or otherwise. This explains the fact (which is something of an accident) that the correspondence principle for the radiation intensity is valid not only in the quasi-classical but in the general quantum case.

§98. **Multipole radiation**

Instead of considering the emission of a photon with a given momentum (i.e. in a given direction), let us now consider the emission of photons with definite values of the angular momentum j. This will also exhibit the deeper quantum-mechanical significance of the dipole approximation.

For the emission of such photons there are rigorous *selection rules* which follow from the law of conservation of angular momentum: the initial angular momentum of the radiating system must be equal to the total angular momentum of the final system and the photon. According to the quantum-mechanical rule for the addition of angular momenta, this means that, if the initial angular momentum of the system is J_i, after the emission of a photon with angular momentum j the angular momentum of the system must have one of the values

$$J_f = J_i + j, \quad J_i + j - 1, \; \ldots, \; |J_i - j|. \tag{98.1}$$

The parities P_i and P_f of the initial and final states of the system must also satisfy a certain condition: the initial parity must be the same as the total parity of the final system and the photon, i.e. $P_f P_{\text{ph}} = P_i$, where P_{ph} is the parity of the photon. Since each parity must be ± 1, this condition may also be written[†]

$$P_i P_f = P_{\text{ph}}. \tag{98.2}$$

The angular momentum of the photon takes integral values from 1 upwards ($j = 0$ is not possible). For any such value the rules (98.1) prohibit the emission of a single photon in a transition of the system between two states with $J = 0$ ($0 \to 0$ transitions). A radiative transition between such states can occur only with the simultaneous emission of two photons having antiparallel angular momenta (but this process appears only in higher approximations of perturbation theory, and therefore has relatively low probability).

For the emission of a photon in the 1^- state (an $E1$ photon in the terminology of §78), the selection rules (98.1) and (98.2) allow transitions only between states with opposite parities and with the following possible changes of the angular momentum J of the radiator:

$$\left.\begin{array}{l} J \to J+1, \quad J, \quad J-1 \quad (\text{for } J \geq 1), \\ \frac{1}{2} \to \frac{3}{2}, \frac{1}{2}, \quad 0 \to 1. \end{array}\right\} \tag{98.3}$$

These are the same as the selection rules for the matrix elements of a polar vector (§§18, 19). The electric dipole moment $\mathbf{d}$ of the system is a polar vector, whose matrix elements determine the probability (97.4). Hence it is clear that the dipole approximation corresponds to the emission of a 1^- photon.

[†] The parity selection rule was first established by O. Laporte (1924).

For the emission of a 1^+ photon ($M1$ photon) the selection rules differ from the electric-dipole case only as regards the parity rule: the initial and final states must have the same parity. This corresponds to the selection rules for the matrix elements of an axial vector. The magnetic dipole moment of the system is an axial vector, and its matrix elements determine in this case the photon emission probability. This is the reason for calling it magnetic dipole radiation.

Similarly, the emission of any Ej photon is determined by the matrix elements of the 2^j-pole electric moment of the system, and that of an Mj photon by those of the 2^j-pole magnetic moment.

§99. Radiation from atoms[†]

The energies of the outer electrons of an atom (which take part in optical radiative transitions) have, as a rough estimate, the order of magnitude $E \sim me^4/\hbar^2$, so that the radiated wavelengths $\lambda \sim \hbar c/E \sim \hbar^2/\alpha me^2$. The dimension of the atom is $a \sim \hbar^2/me^2$. Thus, in the optical spectra of atoms, we generally have the inequality $a/\lambda \sim \alpha \ll 1$. The ratio $v/c \sim \alpha$, where v is the velocity of the optical electrons, has a similar order of magnitude.

Thus, in the optical spectra of atoms, a condition is satisfied which means that the probability of electric dipole radiation (if this is allowed by the selection rules) considerably exceeds the probabilities of higher-order multipole transitions. For this reason it is the $E1$ transitions that are the most important in atomic spectra.[‡]

The selection rules for the total angular momentum and the parity of the electrons in the atom, stated in §98, are rigorous.[||] Together

---

[†] In this section, ordinary units are used.

[‡] Typical values of the dipole transition probability in the optical region of atomic spectra are of the order of 10^8 sec^{-1}.

[||] To avoid misunderstanding, it should be noted that the total angular momentum of the atom (denoted by F in §51) consists of the angular momentum of the electrons and the spin of the nucleus. The most rigorous selection rules must apply to this angular momentum. But, on account of the extremely weak interaction of the electrons with the nuclear spin, the effect of the latter on the electron transition probabilities can be neglected; the selection rules then apply only to the electron characteristics of the state of the atom.

with these, there may be approximate selection rules whose validity depends on certain properties approximately characterising certain categories of atomic states.

Such states are, for example, those based on *LS* coupling (§51). They are characterised by the total angular momentum and also by definite values of the orbital angular momentum L and the spin S of the atom, which in this case are conserved. Since the electric dipole moment is a purely orbital quantity, its operator commutes with the spin operator, i.e. its matrix is diagonal with respect to the number S. For the matrix elements of the dipole moment with respect to the wave functions of the orbital motion of the electrons, there will be selection rules for the number L similar to those for any orbital vector (§18). Thus transitions between states based on *LS* coupling are subject to the additional selection rules:

$$\left.\begin{array}{c} S_f = S_i, \\ L_f = L_i+1, \quad L_i, \quad L_i-1. \end{array}\right\} \tag{99.1}$$

These rules, it may be repeated, are approximate ones, valid only if the spin–orbit interaction, which destroys the separate conservation of the orbital angular momentum and the spin, is neglected.

In the classical theory, the order of magnitude of the magnetic moment of the system (defined as in *Mechanics and Electrodynamics*, (66.2)) is related to that of its dipole moment by $\mu \sim (v/c)d$. A similar relation exists for the quantum-theory atom: the magnetic moment is of the order of the Bohr magneton, $\mu \sim e\hbar/mc$, which differs by a factor α from that of the dipole moment $d \sim ea \sim \hbar^2/me$; since $v/c \sim \alpha$, this gives the above relation between μ and d.

The probability of magnetic dipole ($M1$) radiation is proportional to the square of the magnetic moment, and is therefore less, by a factor of approximately α^2, than the probability of electric dipole radiation at the same frequency. The magnetic radiation is therefore of practical importance only for transitions forbidden by the selection rules for the electric case.

The same is true of the electric quadrupole ($E2$) radiation. The order of magnitude of the electric quadrupole moment of the atom is ea^2.

This contains a further factor a in comparison with the dipole moment $d \sim ea$. Accordingly, the matrix element of the quadrupole radiation transition contains an extra factor $ka \sim a/\lambda$ in comparison with that of the dipole transition; with the orders of magnitude of a and λ stated above, this is again the small factor $\sim \alpha$.

However, the different origin of this factor for $M1$ and $E2$ radiation (from v/c and a/λ respectively) has the result that under certain conditions the $M1$ radiation may be more probable than the $E2$ radiation (if both are allowed by the selection rules, of course). The ratio of their probabilities is

$$\frac{w(E2)}{w(M1)} \sim \frac{(a/\lambda)^2}{(v/c)^2} \sim \left(\frac{a\omega}{v}\right)^2 \sim \left(\frac{\Delta E}{E}\right)^2,$$

where $E \sim v\hbar/a$ is the energy of the atom and $\Delta E = \hbar\omega$ is the change in this energy in the transition. We see that the ratio is of the order of unity if $\Delta E \sim E$, but may be small if $\Delta E \ll E$.

In particular, such a case occurs for transitions between hyperfine-structure components of the same level (the frequencies of such transitions are in the radio range). They cannot occur as electric dipole transitions, since the components of the hyperfine structure differ only in the sum of the electron and nuclear angular momenta, and therefore have the same parity. The $E2$ and $M1$ transitions take place without change of parity. But, because the intervals in the hyperfine structure are relatively very small, the $E2$ radiation has low probability in comparison with $M1$, so that these transitions occur as magnetic dipole transitions.

§100. The infra-red catastrophe

The collision of two charged particles is in general accompanied by the emission of photons (called *bremsstrahlung*). The possible values of the photon frequency form a continuous range from zero up to the total kinetic energy of the relative motion of the colliding particles. Let us consider some of the properties of this radiation in the limiting case of low frequencies.

When the photon energy $\hbar\omega \to 0$, the quantum-mechanical formulae must become the classical ones. Here, of course, the discussion must relate to the calculation of properties of the radiation that are formulated independently of the concept of the photon, such as the total radiation intensity, i.e. the total energy lost by the colliding particles in radiation.

According to the classical theory, the spectral distribution of the bremsstrahlung energy tends, as $\omega \to 0$, to an expression of the form

$$d\mathcal{E} = \text{constant} \times d\omega, \tag{100.1}$$

where the constant is independent of ω; see *Mechanics and Electrodynamics*, §82, Problem 4, which dealt with a non-relativistic collision between two particles having different values of the charge/mass ratio.

Although, according to the previous discussion, this limiting form remains valid in the quantum theory, it has another aspect there. The radiation is described not only by its total energy but by the number of photons emitted. The number of photons with frequencies in the range $d\omega$ is found by dividing $d\mathcal{E}$ by $\hbar\omega$, and in the same limiting case we therefore have

$$dN = \text{constant} \times d\omega/\omega. \tag{100.2}$$

The total number of photons emitted is obtained on integrating $dN/d\omega$ with respect to ω. We see that the integral is (logarithmically) divergent at the lower limit ($\omega = 0$). In the other words, infinitely many photons with infinitesimal energies are emitted. This is referred to as the *infra-red catastrophe*.

We must emphasise that the divergence represents an actual physical situation, and is not related to the fictitious divergences that result from the imperfections of the existing theory. The occurrence of the infra-red divergence is due to the zero mass of the photon, in consequence of which its energy can be arbitrarily small.

Although photons of infinitesimal frequency are not observable in practice, the infra-red divergence is of fundamental significance. Strictly speaking, any collision of charged particles is accompanied by the emission of an infinite number of soft quanta; the probability of a collision with the emission of no photons or of a finite number of photons is zero. In this sense, we can say that a collision of charged particles cannot be strictly elastic. In an exact calculation of the total probability of such collisions, the spectrum of the emitted photons has to be "cut off", and we must agree to regard as "elastic" those cases where photons are emitted with frequencies not exceeding a small but finite limit.

PROBLEM[†]

Determine the cross-section for bremsstrahlung emission when an electron passes through the field of a fixed nucleus with charge $+Ze$. It is assumed that $v \ll c$ but $Ze^2/\hbar v \ll 1$, $Ze^2/\hbar v' \ll 1$, where v and v' are the initial and final velocities of the electron; these latter inequalities are the conditions for the validity of the Born approximation, in which the influence of the field on the wave functions of the electron before and after the collision is neglected.

SOLUTION. In accordance with (97.4), the cross-section for collisions in which a photon with energy $\hbar\omega$ is emitted and the electron acquires a momentum $\mathbf{p}' = m\mathbf{v}'$ in the element of solid angle do' is

$$d\sigma = \frac{4\omega^3}{3\hbar c^3} |\mathbf{d}_{fi}|^2 p'^2 \, dp' \, do'. \tag{1}$$

The additional factor $d^3p = p'^2 \, dp' \, do'$ occurs because the final state (a free electron with momentum $\mathbf{p}'$) belongs to the continuous spectrum. The change from the probability in (97.4) to the cross-section is made by using the appropriate normalisation of the wave function of the initial electron to unit current density:

$$\psi_i = \frac{1}{\sqrt{v}} e^{(i/\hbar)\mathbf{p}\cdot\mathbf{r}}, \tag{2}$$

where $\mathbf{p} = m\mathbf{v}$; cf. (21.6). The wave function of the final electron is a plane wave normalised by the δ-function in momentum space:

$$\psi_f = \frac{1}{(2\pi\hbar)^{3/2}} e^{(i/\hbar)\mathbf{p}'\cdot\mathbf{r}}. \tag{3}$$

The frequency of the emitted photon is related to $\mathbf{p}$ and $\mathbf{p}'$ by the conservation of energy:

$$\hbar\omega = \frac{1}{2m}(p^2 - p'^2). \tag{4}$$

[†] Ordinary units are used.

The matrix element of the electron dipole moment $\mathbf{d} = e\mathbf{r}$ in its motion relative to the centre of the field must, however, be calculated by means of the equation of motion in the field:

$$m\ddot{\mathbf{r}} = \nabla(Ze^2/r)$$

and not directly from the functions (2) and (3). In quantum mechanics, the equation of motion is to be regarded as a relation between the corresponding operators (cf. (21.2)). Taking the matrix elements of these operators, we find

$$m(\ddot{\mathbf{r}})_{fi} = -m\omega^2\mathbf{r}_{fi} = Ze^2\left(\nabla \frac{1}{r}\right)_{fi}.$$

The matrix element on the right, with respect to the functions (2) and (3), becomes the Fourier component

$$\begin{aligned}\left(\nabla \frac{1}{r}\right)_{\mathbf{q}} &= \int \left(\nabla \frac{1}{r}\right) e^{-i\mathbf{q}\cdot\mathbf{r}}\,\mathrm{d}V \\ &= i\mathbf{q}\left(\frac{1}{r}\right)_{\mathbf{q}} \\ &= 4\pi i\mathbf{q}/q^2,\end{aligned}$$

where $\hbar\mathbf{q} = \mathbf{p}' - \mathbf{p}$ and we have used formula (68.6). Then formula (1) gives

$$\mathrm{d}\sigma = \frac{8}{3\pi} Z^2\alpha\left(\frac{e^2}{mc^2}\right)^2 \frac{v'c^2\,\mathrm{d}o'}{v(\mathbf{v}-\mathbf{v}')^2}\frac{\mathrm{d}\omega}{\omega}.$$

To integrate over the directions of $\mathbf{v}'$, we write

$$(\mathbf{v}-\mathbf{v}')^2 = v^2 + v'^2 - 2vv'\cos\theta, \qquad \mathrm{d}o' = 2\pi\sin\theta\,\mathrm{d}\theta,$$

and after the integration with respect to θ we have finally

$$\mathrm{d}\sigma = \frac{16}{3} Z^2\alpha\left(\frac{e^2}{mc^2}\right)^2 \frac{c^2}{v^2}\log\frac{v+v'}{v-v'}\frac{\mathrm{d}\omega}{\omega}.$$

The infra-red catastrophe corresponds to the divergence of this expression as $\omega \to 0$.

§101. Scattering of radiation

The scattering of a photon by an atom consists of the absorption of the initial photon (with momentum $\mathbf{k}$) and the simultaneous emission of another photon ($\mathbf{k}'$). The atom may be left either at its initial energy level or at some other. In the former case the photon frequency is unchanged (*Rayleigh scattering*); in the latter case the frequency changes by

$$\omega' - \omega = E_i - E_f, \qquad (101.1)$$

where E_i and E_f are the initial and final energies of the atom (*Raman* *scattering*). If the atom was initially in the ground state, the frequency can only decrease. In scattering by an excited atom, however, the final level may be either higher or lower, so that Raman scattering may either decrease or increase the frequency.

Since the electromagnetic interaction operator has no matrix elements for transitions in which two photon occupation numbers simultaneously change, the scattering effect appears only in the second approximation of perturbation theory. It must be regarded as taking place via certain intermediate states, which may be of one of two types:

(I) The photon **k** is absorbed and the atom goes from the initial level E_i to one of its other possible levels E_n; in the subsequent transition to the final state, the photon **k**$'$ is emitted;

(II) The photon **k**$'$ is emitted and the atom enters the state E_n, and in the transition to the final state, the photon **k** is absorbed.

According to (36.2), the matrix element for this process is represented by the sum

$$V_{fi} = \sum_n{}' \left(\frac{V'_{fn} V_{ni}}{\mathscr{E}_i - \mathscr{E}_n^{\mathrm{I}}} + \frac{V_{fn} V'_{ni}}{\mathscr{E}_i - \mathscr{E}_n^{\mathrm{II}}}\right). \tag{101.2}$$

Here $\mathscr{E}_i = E_i + \omega$ is the initial energy of the atom+photon system, and $\mathscr{E}_n^{\mathrm{I}}$ and $\mathscr{E}_n^{\mathrm{II}}$ are the energies of the two intermediate states:

$$\mathscr{E}_n^{\mathrm{I}} = E_n, \quad \mathscr{E}_n^{\mathrm{II}} = E_n + \omega + \omega';$$

V_{ni} and V_{fn} are the matrix elements of the transitions with absorption, and V'_{fn} and V'_{ni} those of the transitions with emission; the initial state of the atom is excluded from the summation over n, this being indicated by the prime to the summation sign.

Our problem is to calculate the cross-section of the scattering process. This can be done by means of the same formula (95.14) as was used previously to calculate the probability of spontaneous emission: the only difference is that the "radiator" emitting the photon ω' is now not an isolated atom but a system comprising the atom and the incident photon ω. The cross-section is obtained from the probability by simply dividing the probability by the current density of the photons incident on the atom. The wave function of the photon, normalised to one photon in the volume Ω, corresponds to a current density c/Ω, the product of the velocity c and the photon number density $1/\Omega$. In relativistic units $c = 1$, and the cross-section is therefore calculated from the formula

$$d\sigma = \frac{\Omega^2}{4\pi^2} |V_{fi}|^2 \omega'^2 \, do', \tag{101.3}$$

where do' is the element of solid angle for the directions of the scattered photon.

We shall assume that the wavelengths of the initial and final photons are large compared with the dimensions of the scattering atom. Then the dipole approximation can be used for the matrix elements of all transitions. According to (97.2) and (97.2a),

$$V_{ni} = -i \sqrt{\frac{2\pi\omega}{\Omega}} (\mathbf{e} \cdot \mathbf{d}_{ni}), \quad V'_{fn} = i \sqrt{\frac{2\pi\omega'}{\Omega}} (\mathbf{e}'^* \cdot \mathbf{d}_{fn}),$$

and similarly for V'_{ni} and V_{fn} ($\mathbf{e}$ and $\mathbf{e}'$ being the polarisation vectors of the photons ω and ω').

Substituting these expressions in (101.2) and thence in (101.3), we have the scattering cross-section[†]

$$d\sigma = |A_{fi}|^2 \frac{\omega\omega'^3}{\hbar^2 c^2} \, do', \tag{101.4}$$

where the *scattering amplitude* is

$$\left. \begin{aligned} A_{fi} &= \sum_n \left\{ \frac{(\mathbf{d}_{fn} \cdot \mathbf{e}'^*)(\mathbf{d}_{ni} \cdot \mathbf{e})}{\omega_{ni} - \omega} + \frac{(\mathbf{d}_{fn} \cdot \mathbf{e})(\mathbf{d}_{ni} \cdot \mathbf{e}'^*)}{\omega_{nf} + \omega} \right\}, \\ \hbar\omega_{ni} &= E_n - E_i, \quad \hbar\omega_{nf} = E_n - E_f; \end{aligned} \right\} \tag{101.5}$$

this formula is due to H. A. Kramers and W. Heisenberg (1925). The summation over n is taken over all possible states of the atom,

---

[†] Here and below, ordinary units are used.

including those of the continuous spectrum (the states i and f cannot appear in the sum, since the diagonal matrix elements $\mathbf{d}_{ii}$ and $\mathbf{d}_{ff}$ are zero; see §54).[†]

It is easily seen that the scattering amplitude is zero except for transitions between states of the same parity (including coincident states i and f): the matrix elements of the vector $\mathbf{d}$ are zero except for transitions between states of different parity; the parities of the states i and f must therefore be both opposite to that of a state n which is the same in each term in the sum (101.5), and so they must be equal to each other. This is contrary to the parity selection rule for electric dipole radiation; thus there is an "alternative prohibition", whereby transitions allowed for emission are forbidden for scattering, and vice versa.

When $\omega \to 0$, the scattering amplitude tends to a finite limit. The cross-section for Rayleigh scattering ($\omega' = \omega$) is therefore proportional to ω^4 when ω is small.

In the opposite case, when the frequency ω is large compared with all the frequencies ω_{ni}, ω_{nf} which are important in the sum (101.5) (but of course the wavelength is still much greater than the dimensions of the atom), we must arrive at the formulae of the classical theory. The calculation of the first non-vanishing term in the expansion of (101.5) in powers of $1/\omega$ (which will not be given in detail here) leads to the scattering cross-section

$$d\sigma = Z^2(e^2/mc^2)^2\,|\mathbf{e}'^*\cdot\mathbf{e}|^2\,do', \qquad (101.6)$$

where Z is the number of electrons in the atom. Summation of (101.6) over the polarisations $\mathbf{e}'$ of the scattered photon gives the classical Thomson's formula (*Mechanics and Electrodynamics*, (86.10)).

Let us consider the scattering of radiation by an assembly of N identical atoms situated in a region small compared with the wavelength. The scattering amplitude for such an assembly is equal to the sum of the amplitudes for the individual atoms. It must, however,

---

[†] Formulae (101.4) and (101.5) are not applicable to the case of resonance, when the frequency ω is close to either ω_{ni} or ω_{fn}. In this case (called *resonance fluorescence*), the natural width of the spectral lines has to be taken into account (§102).

be remembered that the wave functions (which are used to calculate the dipole moment matrix elements) for several identical atoms taken together are not simply equal functions. The wave functions are essentially defined only to within an arbitrary phase factor, which is different for each atom. The scattering cross-section has to be averaged over the phase factor of each atom separately.

The scattering amplitude A_{fi} of each atom includes a factor $e^{i(\phi_i-\phi_f)}$, where ϕ_i and ϕ_f are the phases of the wave functions of the initial and final states. For Raman scattering, the states i and f are different, and this factor is not equal to unity. In the squared modulus $|\Sigma A_{fi}|^2$, which determines the scattering cross-section (the sum being over all N atoms), the products of terms pertaining to different atoms will include phase factors which vanish on averaging over the phases of the atoms, and only the squared modulus of each term remains. This means that the total cross-section for scattering by N atoms is found by taking N times the cross-section for scattering by one atom: the scattering cross-sections are added, not the scattering amplitudes. In this case the scattering is said to be *incoherent*.

If, however, the initial and final states of the atom are the same, then the factors $e^{i(\phi_i-\phi_f)} = 1$. The scattering amplitude is then N times the scattering amplitude for one atom, and the scattering cross-section consequently differs by a factor N^2. In this case the scattering is said to be *coherent*.

Coherent scattering is certainly of the Rayleigh type, but the converse statement is not necessarily true. Rayleigh scattering is entirely coherent only if the scattering atom is at a non-degenerate energy level. If the energy level is degenerate, there will also be incoherent Rayleigh scattering arising from transitions of the atom between various mutually degenerate states. It must be emphasised that the incoherence of Rayleigh scattering is a purely quantum effect; in the classical theory, scattering without change of frequency is necessarily coherent, and the concept of coherent scattering has been defined in this way in *Mechanics and Electrodynamics*, §86.

§102. Natural width of spectral lines

So far, in the study of emission and scattering of radiation, we have regarded all the levels of the system (an atom, say) as being strictly discrete. But in fact excited levels have a certain probability of emission, and therefore a finite lifetime. This has the result that the levels become quasi-discrete, with a certain small but finite width; they can be written in the form $E-\frac{1}{2}i\Gamma$, where Γ ($\Gamma/\hbar$ in ordinary units) is the probability (per unit time) of all possible processes of "decay" of the state concerned (§38).[†]

Let us consider how this situation affects the process of emission. It is evident that, because of the finite width of the level, the emitted radiation will not be strictly monochromatic: its frequencies will be spread over a range $\Delta\omega \sim \Gamma$. But, in order to measure the frequency distribution of the photons with this accuracy, the time needed is $T \gg 1/\Delta\omega \sim 1/\Gamma$. During this time the level will almost certainly decay by emission. We therefore have to deal with the total probability of emission of a photon of a given frequency, not with the probability per unit time. We shall calculate this total probability, first of all, for a transition of an atom from some excited level $E_i-\frac{1}{2}i\Gamma_i$ to the ground level (E_f), which has an infinite lifetime and is therefore strictly discrete. To simplify the analysis, we shall assume that this transition is the only means of emission from the excited level concerned.

Let us return to the derivation in §35 of the formula (35.6) for the transition probability (used in §95 to calculate the emission probability). The function $a_{fi}(t)$ was considered for large values of t, and the ratio $|a_{fi}|^2/t$ gave the required transition probability per unit time. We can now refine the significance of this procedure: it relates to times short compared with the lifetime of the excited level, "large values of t" here meaning times long compared with the period $1/(E_i-E_f)$ but still small compared with $1/\Gamma$. For this reason it was possible to neglect the finite width of the level. Now that we have to consider times comparable with $1/\Gamma$, the width of the excited level cannot be neglected.

[†] The radiation width of the levels is in practice very small. For example, a decay probability $w \sim 10^8-10^9\ \text{sec}^{-1}$ corresponds to a width $\Gamma \sim 10^{-8}-10^{-7}$ eV.

In the emission problem, the atom+photons system is concerned; accordingly, the transition frequency ω_{fi} in (35.2) becomes $E_f+\omega-E_i$. Writing the initial level in the form $E_i-\frac{1}{2}i\Gamma_i$, we have

$$a_{fi}(t) = V_{fi}\frac{1-\exp\{i(E_f+\omega-E_i)t-\frac{1}{2}\Gamma_i t\}}{E_f-E_i+\omega+\frac{1}{2}i\Gamma_i}. \qquad (102.1)$$

The required transition probability (over all time) is given by the limit of $|a_{fi}(t)|^2$ as $t\to\infty$. For the emission of a photon with frequency in the range $d\omega$ and direction in the solid angle do, it is

$$dW = |a_{fi}(\infty)|^2\frac{\Omega\omega^2\,d\omega\,do}{(2\pi)^3}, \qquad (102.2)$$

where Ω is, as in (95.13), the normalisation volume for the photon wave function. Substituting (102.1), we obtain

$$dW = \frac{\Omega\omega^2\,do}{(2\pi)^3}|V_{fi}|^2\frac{d\omega}{[\omega-(E_i-E_f)]^2+\frac{1}{4}\Gamma_i^2}.$$

To find the spectral distribution of the emission probability, we integrate this over the directions of the photon. According to (95.14),

$$\int\frac{\Omega\omega^2}{(2\pi)^3}|V_{fi}|^2\,do = \frac{w}{2\pi},$$

where w is the total emission probability per unit time, which by definition is equal to Γ_i. Thus we have, finally,

$$dW = \frac{\Gamma_i}{2\pi}\frac{d\omega}{[\omega-(E_i-E_f)^2]+\frac{1}{4}\Gamma_i^2}. \qquad (102.3)$$

The integration of this expression over all frequencies from $-\infty$ to $+\infty$ gives unity, in accordance with the fact that the atom will certainly emit a photon of some frequency during an infinite time.

Formula (102.3) determines what is called the shape of the spectral line, i.e. the distribution of the intensity over the width of a line. The shape given by (102.3) is that for an isolated atom, and is called the *natural* shape of the line.†

---

† As distinct from the broadening caused by the interaction of the atom with other atoms (collision broadening) or by the presence of atoms in the source which move with various velocities (Doppler broadening).

CHAPTER 16

FEYNMAN DIAGRAMS

§103. The scattering matrix

It has already been mentioned in §75 that a typically stated problem in relativistic quantum theory is to determine the probability amplitudes of various scattering processes (transitions between different states of a system of free particles). This problem may be regarded as now solved in principle within the terms of quantum electrodynamics, i.e. for processes governed by the electromagnetic interaction. The weakness of this interaction (which is expressed by the smallness of the fine-structure constant α) enables us to treat such processes by means of perturbation theory. In its usual form (as applied in non-relativistic quantum mechanics), the formalism of perturbation theory has the disadvantage that it does not display explicitly the requirements of relativistic invariance. This disadvantage is removed in a consistent relativistic perturbation theory due to R. P. Feynman (1948). The formalism of this theory allows a very great simplification of calculations which might indeed be impracticable with the ordinary perturbation theory. It also make possible an unambiguous elimination of the divergences (already mentioned in §75) that occur in the calculations.[†]

We shall first of all show how the most general expression for the scattering amplitudes of arbitrary processes can be derived.

[†] The discussion in this chapter is intended to give only an understanding of the basic ideas of the theory, and of the origin and significance of the concepts and quantities that appear in it. The necessary calculations are therefore not given in their entirety; only their general outline is indicated, in order to elucidate the underlying ideas.

With a view to the second-quantised description of the system of particles, we shall use a wave function of the system in which the independent variables are the occupation numbers of the states of free particles; let this function be denoted by Φ to distinguish it clearly from the ordinary coordinate wave functions. The Hamiltonian of the system may be written as $\hat{H} = \hat{H}_0 + \hat{V}$, where $\hat{H}_0$ is the Hamiltonian of the free particles and $\hat{V}$ is the electromagnetic interaction operator. The function Φ satisfies the wave equation

$$i\, \partial\Phi/\partial t = (\hat{H}_0 + \hat{V})\Phi, \tag{103.1}$$

where the ordinary (Schrödinger) representation of the operators and wave functions is used: the operators are independent of time, and the time variation of the system is described by the time dependence of the wave function.

It has been mentioned in §76 that another treatment of the formalism of quantum mechanics is also possible, in which the explicit time dependence is transferred from the wave functions to the operators; in this (Heisenberg) representation, the wave functions are independent of time. For the problem at present under consideration, however, the most suitable representation is an "intermediate" one, in which not the whole time dependence is transferred to the operators, but only the part that corresponds to the state of a system of free particles. In this representation (called the *interaction representation*), therefore, the wave function is time-dependent, but this dependence is entirely due to the action of the perturbation, i.e. corresponds just to the relevant scattering processes, which are due to the interaction of the particles.

Accordingly, the wave equation for the function Φ in the interaction representation is

$$i\, \partial\Phi/\partial t = \hat{V}(t)\Phi, \tag{103.2}$$

which differs from (103.1) by the absence of $\hat{H}_0$ on the right-hand side. In the operator $\hat{V}$, the argument t is shown in order to emphasise that in this representation it is time-dependent, unlike the time-independent Schrödinger operator $\hat{V}$ in (103.1).

If $\Phi(t)$ and $\Phi(t+\delta t)$ are the values of Φ at two successive instants, (103.2) shows that

$$\Phi(t+\delta t) = [1-i\,\delta t.\hat{V}(t)]\,\Phi(t),$$

or, to the same accuracy, $\Phi(t+\delta t) = \exp\,[-i\,\delta t.\hat{V}(t)]\Phi(t)$. Applying this formula to successive time intervals δt_n from $t = -\infty$ to $t = +\infty$, we can express the final value $\Phi(+\infty)$ in terms of the initial value $\Phi(-\infty)$. Denoting by $\hat{S}$ the operator that relates these values, we have $\Phi(+\infty) = \hat{S}\Phi(-\infty)$, where

$$\hat{S} = \prod_n \exp\,[-i\,\delta t_n.\hat{V}(t_n)], \tag{103.3}$$

and the symbol Π denotes the limit of the product over all the intervals δt_n. If $V(t)$ were an ordinary function, this limit would reduce simply to

$$\exp\left[-i\sum_n \delta t_n.V(t_n)\right] = \exp\left(-i\int_{-\infty}^{\infty} V(t)\,\mathrm{d}t\right),$$

but this result depends on the commutativity of the factors pertaining to different instants, which is assumed in changing from the product in (103.3) to the summation in the exponent. For the operator $\hat{V}(t)$ there is in general no such commutativity and the reduction to an ordinary integral is not possible.

We can write (103.3) in the symbolic form

$$\hat{S} = \hat{T}\exp\left\{-i\int_{-\infty}^{\infty}\hat{V}(t)\,\mathrm{d}t\right\}, \tag{103.4}$$

where $\hat{T}$ is the *chronological operator*, implying a certain "chronological" sequence of time instants in the successive factors of the product (103.3). This notation in itself is, of course, no more than symbolic, but it allows a simple derivation of the series expansion of $\hat{S}$ in powers of the perturbation:

$$\hat{S} = \sum_{k=0}^{\infty}\frac{(-i)^k}{k!}\int_{-\infty}^{\infty}\mathrm{d}t_1\int_{-\infty}^{\infty}\mathrm{d}t_2\ldots\int_{-\infty}^{\infty}\mathrm{d}t_k\hat{T}\{\hat{V}(t_1)\hat{V}(t_2)\ldots\hat{V}(t_k)\}. \tag{103.5}$$

Here, in each term, the kth power of the integral is written as a k-fold integral, and the operator $\hat{T}$ signifies that in each range of values of the variables $t_1, t_2, \ldots, t_k$ the factors $\hat{V}(t_1), \hat{V}(t_2), \ldots, \hat{V}(t_k)$ must be put in chronological order, with the value of t increasing from right to left. Since the time-ordering operation now relates simply to a product (and not to an exponential as in (103.4)), the expression for each term in the sum (103.5) is not merely symbolic, but has real significance.

It is clear from the definition of the operator $\hat{S}$ that, if the system was in the state Φ_i (an assembly of free particles) before the collision, the probability amplitude for its transition to the state Φ_f (another assembly of free particles) is the matrix element S_{fi}. For, according to the definition of the matrix elements of an operator, the function $\Phi(\infty) = \hat{S}\Phi_i$ can be written as the expansion

$$\Phi(\infty) = \sum_f S_{fi}\Phi_f$$

(cf. (11.11)); $|S_{fi}|^2$ is therefore the probability that the system is in the final state Φ_f as $t \to \infty$ (i.e. after the interaction). The operator $\hat{S}$ is called the *scattering operator*, and its matrix elements form the *scattering matrix* or *S-matrix* (a term already mentioned in §75). The non-diagonal ($i \neq f$) elements of this matrix are the amplitudes of the scattering processes $i \to f$.[†]

To arrive at a fully specific meaning for formula (103.5), we have still to establish the general form of the interaction operator $\hat{V}(t)$, so as to include all possible electrodynamic processes. This is easily done by a direct generalisation of the formulae already given in §95, where second quantisation was applied only to the electromagnetic field represented by the operator $\hat{\mathbf{A}}$ in (95.1). We now have to carry out this process for the electron–positron field also. This is done by simply substituting the appropriate Ψ-operators for the electron wave functions in the matrix elements (95.5), (95.6). Thus we obtain the expression

$$\hat{V}(t) = -e\int \hat{\mathbf{j}}\cdot\hat{\mathbf{A}}\, d^3x, \tag{103.6}$$

[†] The derivation of the rules of relativistic perturbation theory by means of the expansion (103.5) is due to F. J. Dyson.

where $\hat{\mathbf{j}} = \Psi^* \boldsymbol{\alpha} \Psi$ is the second-quantised particle current density operator and $d^3x = dx\, dy\, dz$ is the volume element.

In (103.6) we have the three-dimensional vectors $\mathbf{j}$ and $\hat{\mathbf{A}}$, on account of the particular choice of gauge used hitherto for the field potentials, namely that in which the scalar potential is zero. In order to derive relativistically invariant expressions, we must now use the four-dimensional notation

$$\hat{V}(t) = e \int \hat{j}^\mu \hat{A}_\mu \, d^3x, \tag{103.7}$$

where $\hat{j}^\mu = \hat{\overline{\Psi}} \gamma^\mu \hat{\Psi}$ is the current density four-vector operator and $\hat{A}_\mu$ is the four-potential operator without any predetermined choice of gauge; when $\hat{A}^\mu = (0, \hat{\mathbf{A}})$, (103.7) becomes (103.6). The form of the operator $\hat{A}^\mu$ differs from (76.15) only in that the photon polarisation vector $\mathbf{e}$ is replaced by a unit four-vector e^μ (which reduces to $(0, \mathbf{e})$ only in a particular gauge):[†]

$$\hat{A}^\mu = \sum_{\mathbf{k}} \sqrt{\frac{2\pi}{\Omega\omega}} \left(\hat{c}_{\mathbf{k}} e^\mu e^{-i(kx)} + \hat{c}_{\mathbf{k}}^+ e^{\mu *} e^{i(kx)} \right). \tag{103.8}$$

The Ψ-operators are expressed in terms of the electron and positron creation and annihilation operators by formulae (85.3). They may be written

$$\hat{\Psi} = \sum_{\mathbf{p}} (\hat{a}_{\mathbf{p}} \Psi_p + \hat{b}_{\mathbf{p}}^+ \Psi_{-p}), \qquad \hat{\overline{\Psi}} = \sum_{\mathbf{p}} (\hat{a}_{\mathbf{p}}^+ \overline{\Psi}_p + \hat{b}_{\mathbf{p}} \overline{\Psi}_{-p}), \tag{103.9}$$

where the functions Ψ_p are plane waves with four-momenta p:

$$\Psi_p = (1/\sqrt{\Omega})\, u(p) e^{-i(px)}. \tag{103.10}$$

[†] For brevity, the indices denoting the polarisation of the particles are everywhere omitted. In this chapter, we shall frequently use the conventional notation of light-face letters for four-vectors without the indices $\mu, \nu, \ldots$ that label their components. For example, x and p denote the four-vectors $x^\mu = (t, \mathbf{r})$ and $p^\mu = (\varepsilon, \mathbf{p})$. The scalar products of four-vectors are likewise written without indices. For instance, $(px) \equiv p_\mu x^\mu = \varepsilon t - \mathbf{p}.\mathbf{r}$; the equation $p_\mu p^\mu = m^2$ for the four-momentum of a particle with mass m is written $p^2 = m^2$, the equation $k_\mu k^\mu = 0$ for the four-momentum of a photon is written $k^2 = 0$, and so on. This notation is often used in recent literature. It is a compromise between the limited resources of the alphabet and the demands of physics, and means, of course, that the reader must be more than usually attentive.

It should be noted that the time dependence of the operators (103.8) and (103.9), and therefore that of the interaction operator (103.7), have been transferred to them from the wave functions of free motion of particles (plane waves). Thus these operators are in fact in the required interaction representation.

§104. Feynman diagrams

The procedure for calculating the elements of the scattering matrix can be illustrated by means of some specific examples. Let us consider processes occurring in the second approximation of perturbation theory. These correspond to the second-order term ($k = 2$) in (103.5). Substituting for $\hat{V}$, we can write this term as

$$\hat{S}^{(2)} = -\tfrac{1}{2}e^2 \iint \mathrm{d}^4x\, \mathrm{d}^4x' \hat{T}\{\hat{j}^\mu(x)\, \hat{A}_\mu(x) \hat{j}^\nu(x')\, \hat{A}_\nu(x')\}, \qquad (104.1)$$

where $\mathrm{d}^4x = \mathrm{d}t\, \mathrm{d}^3x$ is the four-volume element. This formula is relativistically invariant: the products $(\hat{j}\hat{A})$ are four-scalars, and the integration over the four-volume is a scalar operation.[†]

As a first example, let us take the elastic scattering of two electrons: in the initial state there are two electrons with four-momenta p_1 and p_2, and in the final state there are two electrons with other four-momenta p_3 and p_4. Since the photon and electron operators act on different variables (the photon and electron occupation numbers), their matrix elements can be calculated independently. In the present case, there are no photons in the initial and final states, and so the appropriate matrix element with respect to the photon operators $\hat{A}_\mu(x)\, \hat{A}_\nu(x')$ is the diagonal element $\langle 0 | \ldots | 0 \rangle$, where the symbol $| 0 \rangle$ denotes the state of the electromagnetic field without photons (the *photon vacuum* state). This matrix element is a certain function of the four-coordinates x and x'. Because of the homogeneity of space and time, this function can depend only on the space and time intervals $\mathbf{r}-\mathbf{r}'$ and $t-t'$, i.e. only on the difference $x-x'$, and not on the values of x and x' sepa-

---

[†] We shall not pause here to prove that the relativistic invariance is also unaffected by the operation of time-ordering.

rately. Thus we arrive one of the fundamental new concepts of this theory, the *photon propagation function* or *photon propagator*, defined as

$$D_{\mu\nu}(x-x') = \begin{Bmatrix} i\langle 0 \,|\, A_\mu(x)\, A_\nu(x') \,|\, 0\rangle & \text{for } t' < t, \\ i\langle 0 \,|\, A_\nu(x')\, A_\mu(x) \,|\, 0\rangle & \text{for } t < t'; \end{Bmatrix} \tag{104.2}$$

the different order of the factors for $t' < t$ and $t < t'$ is due to the action of the operator $\hat{T}$ in (104.1).

Let us next consider the electron operators in (104.1). Each of the two current operators is a product $\hat{j} = \hat{\overline{\Psi}}\gamma\hat{\Psi}$, and each of the Ψ-operators is given by the sum (103.9). Hence the product $\hat{j}^\mu(x)\hat{j}^\nu(x')$ is a sum of terms, each containing a product of four of the operators $\hat{a}_{\mathbf{p}}, \hat{a}^+_{\mathbf{p}}, \hat{b}_{\mathbf{p}}, \hat{b}^+_{\mathbf{p}}$. A non-zero contribution to the matrix element in question comes from the terms in which the operators annihilate the initial electrons p_1, p_2 and create the final electrons p_3 and p_4. Thus the terms required are those containing a product of the operators $\hat{a}_{\mathbf{p}_1}$, $\hat{a}_{\mathbf{p}_2}$, $\hat{a}^+_{\mathbf{p}_3}$, $\hat{a}^+_{\mathbf{p}_4}$. The calculation based on this gives the result

$$\begin{aligned} S_{fi} = ie^2 \iint \mathrm{d}^4x\, \mathrm{d}^4x'\, D_{\mu\nu}(x-x')\{&(\overline{\Psi}_4\gamma^\mu\Psi_2)\,(\overline{\Psi}{}'_3\gamma^\nu\Psi'_1) \\ &-(\overline{\Psi}_4\gamma^\mu\Psi_1)\,(\overline{\Psi}{}'_3\gamma^\nu\Psi'_2)\}, \end{aligned} \tag{104.3}$$

where $\Psi_1 = \Psi_{p_1}(x)$, $\Psi'_1 = \Psi_{p_1}(x')$, etc.

The electron wave functions are the plane waves (103.10). Hence, for example, the first term in the braces in (104.3) contains the exponential factor

$$\exp\{-i[(p_2-p_4)x]-i[(p_1-p_3)x']\}.$$

From the conservation of four-momentum in the collision $p_1+p_2 = p_3+p_4$, or $p_2-p_4 = p_3-p_1$. This factor thus becomes

$$\exp\{i[(p_4-p_2)(x-x')]\}$$

and the integration over $\mathrm{d}^4(x-x')$ in (104.3) corresponds to taking the component corresponding to the four-momentum $k = p_4-p_2$ in the expansion of the function $D_{\mu\nu}(x-x')$ as a four-dimensional Fourier integral corresponding to $k = p_4 - p_2$. The function

$$D_{\mu\nu}(k) = \int D_{\mu\nu}(x-x')e^{i[k(x-x')]}d^4(x-x') \qquad (104.4)$$

is called the *photon propagator in the momentum representation.*

The second term in (104.3) is transformed similarly, and the result is

$$S_{fi} \sim e^2(\bar{u}_4\gamma^\mu u_2)\,D_{\mu\nu}(k)\,(\bar{u}_3\gamma^\nu u_1) - e^2(\bar{u}_4\gamma^\mu u_1)\,D_{\mu\nu}(k')\,(\bar{u}_3\gamma^\nu u_2), \qquad (104.5)$$

where $k = p_4 - p_2$, $k' = p_4 - p_1$.[†] The two terms in this scattering amplitude can be symbolically represented by what are called *Feynman diagrams* (Fig. 14). Each point of intersection of lines (a *vertex* of the diagram) has a corresponding factor $e\gamma^\mu$. The "incoming" continuous lines towards a vertex represent the initial electrons, which are associated with the factors u, the bispinor amplitudes of the corresponding electron states. The "outgoing" continuous lines leaving a vertex are the final electrons, and correspond to the factor $\bar{u}$. When the diagram is "read", these factors are written from left to right in the order of movement along the continuous lines against the direction of the arrows. The two vertices are joined by a broken line which represents a *virtual* (intermediate) *photon* "emitted" at one vertex and "absorbed" at the other, and corresponds to the factor $D_{\mu\nu}(k)$. The four-momentum of the virtual photon (k or k') is determined by the "conservation of four-momentum" at the vertex; the total mo-

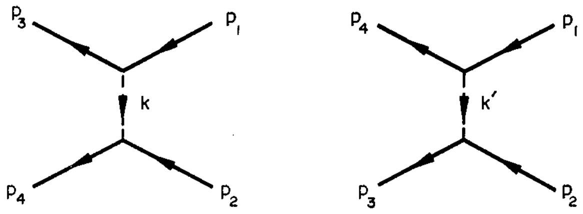

Fig. 14

[†] We are here concerned only with the mathematical structure of the S-matrix elements, and therefore omit all irrelevant common factors. We shall also pass over the question of how $|S_{fi}|^2$ is converted into an observable quantity, the scattering cross-section.

menta of the incoming and outgoing lines are equal. The lines corresponding to the initial and final particles are called *external lines* or *free ends* of the diagram. The two diagrams in Fig. 14 differ by the interchange of two such lines.

The square of the virtual photon four-momentum, $k^2 \equiv k_\mu k^\mu$, is *not* zero, as it would have to be for a real photon. Moreover, the description of the process (in accordance with the form of the diagrams) as the emission of a virtual photon followed by its absorption has, of course, no literal significance; it is merely a convenient way of describing in words the structure of the expressions that occur in the scattering amplitude.

Let us now consider the scattering of an electron and a positron. Their initial and final four-momenta will be denoted by p_-, p_+; p'_-, p'_+. The appropriate changes in the diagrams are clear from the structure of the Ψ-operators (103.9): the positron creation and annihilation operators appear in these expressions together with the electron annihilation and creation operators respectively, and with coefficients $\bar{u}(-p)$ and $u(-p)$ instead of $u(p)$ and $\bar{u}(p)$. Thus the diagrams in Fig. 14 are replaced by those in Fig. 15. The rules for constructing the diagrams are modified only as regards the positrons. As

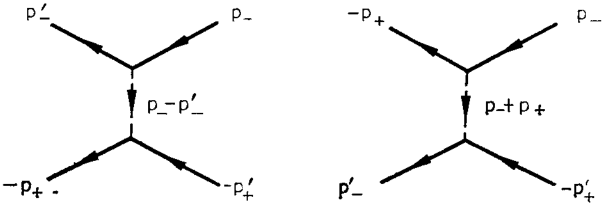

FIG. 15

before, the incoming continuous lines are associated with the factors u, and the outgoing ones with $\bar{u}$. But the incoming lines now correspond to the final positrons and the outgoing ones to the initial positrons, the four-momenta of all the positrons being taken with reversed sign. The two diagrams in Fig. 15 differ: one has the same structure as those in Fig. 14, with the initial and final electron lines meeting at one vertex and the positron lines meeting at the other (a "scattering" type diagram), but in the second diagram initial or final electron and positron lines meet at each vertex; the pair is, as it were, annihilated at the upper vertex with the emission of a virtual photon and produced from this photon at the lower vertex (an "annihilation" type diagram).

Let us now consider another second-order effect, the scattering of a photon by an electron (the *Compton effect*). In the initial state let the photon and the electron have four-momenta k_1 and p_1, and in the final state k_2 and p_2.

In the corresponding element of the S-matrix, the operators $\hat{A}_\mu(x)$ $\hat{A}_\nu(x')$ in (104.1) annihilate the photon k_1 and create the photon k_2 (by virtue of the operators $\hat{c}_{\mathbf{k}_1}$ and $\hat{c}^+_{\mathbf{k}_2}$ in them). The annihilation of the electron p_1 and the creation of the electron p_2 is due to one of the two pairs of operators $\hat{\Psi}$ and $\hat{\bar{\Psi}}$ (by virtue of the operators $\hat{a}_{\mathbf{p}_1}$ and $\hat{a}^+_{\mathbf{p}_2}$). As regards the second pair of Ψ-operators in (104.1), there then remains the diagonal matrix element $\langle 0|\ldots|0\rangle$, where the symbol $|0\rangle$ now denotes the *electron–positron vacuum* state—a field without particles. Thus we arrive at a second fundamental concept of the theory, the *electron propagation function* or *electron propagator*, defined as

$$G_{ik}(x-x') = \begin{cases} -i\langle 0\,|\,\hat{\Psi}_i(x)\,\hat{\bar{\Psi}}_k(x')\,|\,0\rangle & \text{for} \quad t' < t, \\ i\langle 0\,|\,\hat{\bar{\Psi}}_k(x')\,\hat{\Psi}_i(x)\,|\,0\rangle & \text{for} \quad t < t'. \end{cases} \tag{104.6}$$

Here i and k are bispinor indices, so that G_{ik} is a bispinor of rank two.

The scattering amplitude is found to be

$$S_{fi} \sim e^2\bar{u}_2(e_2^*\gamma)\,G(p)\,(e_1\gamma)u_1 + e^2\bar{u}_2(e_1\gamma)\,G(p')\,(e_2^*\gamma)u_1, \tag{104.7}$$

where $p = p_1+k_1$, $p' = p_1-k_2$; e_1 and e_2 are the polarisation four-vectors of the initial and final photons[†]; $G(p)$ and $G(p')$ are the electron propagators in the momentum representation.

[†] The notation for the polarisation four-vectors should not be confused with the charge e whose square appears as a coefficient in (104.7).

The two terms in this expression are represented by the Feynman diagrams in Fig. 16. The broken external lines correspond to real photons; the incoming lines (initial photons) are associated with the (four-vector) factor e_1, and the outgoing lines (final photons) with e_2^*. The continuous internal line joining the two vertices corresponds to a virtual electron, with the factor $G(p)$. The four-momentum of this virtual electron (p or p') is determined by the conservation of four-momentum at the vertices; its square in *not* equal to m^2 as it would have to be for a real electron.

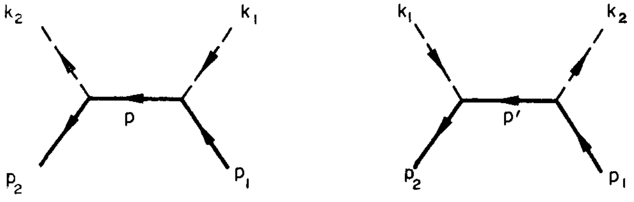

FIG. 16

In the same way as the electron–positron scattering diagrams were obtained by changing the nature of the external electron lines in Fig. 14, we can get from Fig. 16 the diagrams that describe another process, the annihilation of an electron p_- and a positron p_+ to form two photons k_1 and k_2 (Fig. 17).

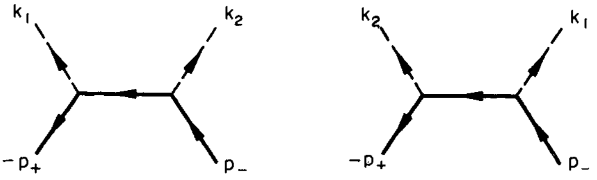

FIG. 17

The rules given here for specific examples form the basis of what is called the *diagram technique*, which can be used to construct the amplitudes of various electrodynamic processes. The amplitude of a scattering process that occurs in the nth approximation of perturbation theory is represented by the set of all diagrams containing n vertices and the same number of external lines as there are initial and final particles concerned in the process. Three lines meet at each vertex: one photon line and two electron lines (one incoming and one outgoing).

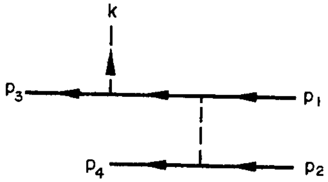

FIG. 18

For example, the three-vertex diagram in Fig. 18 is one of the eight diagrams that correspond (in the third order of perturbation theory) to the emission of a photon k in the collision of electrons with four-momenta p_1 and p_2 (and p_3, p_4 after the collision). In this diagram, the photon k is emitted by one of the final electrons; in the other diagrams, it is emitted by the other electrons (and also p_3 and p_4 may be interchanged).

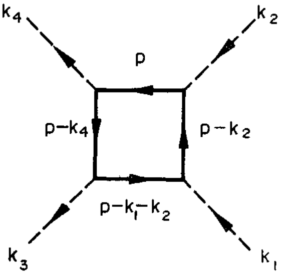

FIG. 19

The fourth-order diagram in Fig. 19 is one of six which describe photon–photon scattering; the other five differ from it by interchanges of the four photon lines.† In comparison with the earlier diagrams,

---

† Photon–photon scattering is a specifically quantum-electrodynamic process; in classical electrodynamics it does not occur, because Maxwell's equations are linear. The existence of this effect indicates that quantum phenomena bring about small non-linear terms in Maxwell's equations.

Fig. 19 has the feature that the conservation of four-momentum at its vertices (for given initial k_1, k_2 and final k_3, k_4) does not uniquely determine the four-momenta of the virtual electrons (the internal continuous lines in the diagram); one of them can be assigned an arbitrary value p. Then the expression obtained from the diagram must also be integrated over all values of the components of the four-vector p.

The concept of the propagators plays a fundamental role in the formalism of quantum electrodynamics. They need to be calculated once and for all, in order to determine the actual scattering amplitudes. This calculation is based on the following important mathematical property of the propagators.

The operator $\hat{\Psi}(x)$ satisfies Dirac's equation $[(\hat{p}\gamma)-m]\hat{\Psi}(x) = 0$, since each of the wave functions Ψ_p in the expansion (103.9) does so. Hence the function $G(x-x')$ (in whose definition $\hat{\Psi}(x)$ appears according to (104.6)) is reduced to zero by the operator $(\gamma\hat{p})-m$ at all points x except those where $t = t'$. The reason is that, according to the definition (104.6), the function $G(x-x')$ tends to different limits as t tends to t' from above and from below ($t \to t'+0$ and $t \to t'-0$). A calculation of the difference between the limits leads to the simple result that G has a discontinuity at $t = t'$ given by

$$\begin{aligned}\Delta G &\equiv [G]_{t\to t'+0}-[G]_{t\to t'-0}\\ &= -i\gamma^0\,\delta(\mathbf{r}-\mathbf{r}').\end{aligned}$$

But, if the function $G(t-t', \mathbf{r}-\mathbf{r}')$ has a discontinuity ΔG at $t-t' = 0$, this means that its derivative $\partial G/\partial t$ has a δ-function term $\Delta G\cdot\delta(t-t')$.[†] In the operator $(\gamma\hat{p})-m$, the time derivative appears in the form $i\gamma^0\partial/\partial t$. Thus we have, finally,

$$[(\gamma\hat{p})-m]\,G(x-x') = \delta^{(4)}(x-x'),$$

where the symbol $\delta^{(4)}$ denotes the product of four δ-functions of the

---

[†] For, on integrating the derivative $\partial G/\partial t$ over a small interval of time t around t', we must obtain the difference between the values of G on either side of $t = t'$; since the integration of the δ-function gives unity, the result is ΔG, as it should be.

four components of the four-vector argument: $\delta^4(x-x') = \delta(t-t')\delta(\mathbf{r}-\mathbf{r}')$.

Thus the function $G(x-x')$ satisfies an inhomogeneous differential equation consisting of Dirac's equation with a δ-function on the right-hand side. In mathematical physics, such a function is called a Green's function of the corresponding homogeneous equation, in this case Dirac's equation. The electron propagator is therefore also frequently called the *electron Green's function.*

Similarly, the photon propagator is the Green's function of the wave equation satisfied by the electromagnetic field potentials, and so it is customarily known also as the *photon Green's function.*

§105. Radiative corrections

The diagram technique allows us, in principle, to calculate not only the scattering amplitudes in the first non-vanishing approximation of perturbation theory but also the corrections to them in higher approximations. These are called *radiative corrections.*

In the calculation of such corrections, difficulties usually arise because of the occurrence of divergent integrals. This demonstrates the logical incompleteness of the existing quantum electrodynamics. In this theory, however, it is possible to establish certain rules which allow an unambiguous "subtraction of infinities", and thus to obtain finite values for all quantities that have an observable physical significance. These rules are based on obvious physical requirements, namely that the photon mass must be zero, and that the electron mass and charge must be equal to their observed values. The assignment of predetermined values to divergent expressions on the basis of physical requirements is called *renormalisation* of the corresponding quantities.

The diagrams representing radiative corrections to the scattering amplitudes are obtained from the basic diagrams by adding new vertices, leaving the number of external lines unchanged. For example, a virtual photon line in a diagram may be given an "electron loop" with two new vertices (Fig. 20a). The four-vector p remains arbitrary, and integration must be performed with respect to it; this integral is divergent, and needs to be renormalised. This diagram can be intuitively described as representing the creation from the vacuum of a virtual electron–positron pair (with four-momenta p and $k-p$) by a virtual photon k, followed by annihilation of the pair to form the same photon as before. Consequently, the radiative corrections pertaining to diagrams of the type shown in Fig. 20a are referred to as the *vacuum polarisation* effect. This leads, in particular, to a distortion of the Coulomb field near a charged particle.[†]

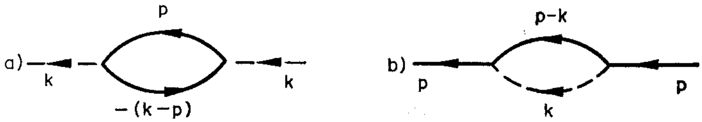

FIG. 20

Similarly, by adding two further vertices we can modify a virtual electron line as shown in Fig. 20b. The virtual electron p can be regarded as emitting a virtual photon and then absorbing it again.

The interaction of an electron with a photon is represented in Feynman diagrams by a vertex at which a photon line k meets electron lines p_1, p_2 (Fig. 21a). The more complicated "diagram section" (Fig. 21b) represents the radiative correction to a simple vertex. This correction leads, in particular, to an important result: the magnetic moment μ of the electron is no longer strictly equal to the value (93.9)

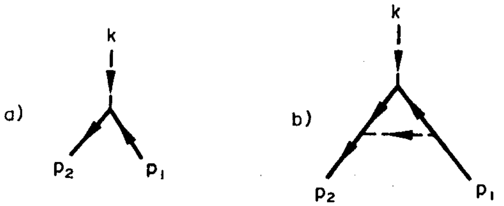

FIG. 21

[†] These distortions extend to distances $\sim \hbar/mc$, where m is the electron mass.

given by Dirac's equation. When the radiative correction is taken into account, μ is (in ordinary units)

$$\mu = \frac{e\hbar}{2mc}\left(1+\frac{\alpha}{2\pi}\right),$$

where α is the fine-structure constant, a formula first derived by J. Schwinger (1949).

§106. Radiative shift of atomic levels

One of the most interesting effects of the radiative corrections is the shift of the atomic energy levels called the *Lamb shift*. This leads, in particular, to the removal of the degeneracy of the hydrogen atom levels that remains even when Dirac's equation is applied (§94). It is impossible to give here a complete analysis of this correction; what follows is a simple derivation in terms of non-relativistic theory. Although this derivation is not entirely consistent, it will serve to illustrate the origin of the radiative corrections.†

The operator of the interaction between an electron system (say a hydrogen atom) and a photon field has no diagonal matrix elements (§95). In the first approximation of perturbation theory, therefore, this interaction gives no correction to the energy levels of the atom. There is a correction in the second approximation, however. According to the general formula (32.10), the second-order correction to the energy levels is given by the non-diagonal matrix elements of the perturbation that correspond to transitions from a specified state to intermediate states. In the present case, we are concerned with states of the system consisting of the atom and the photon field; the initial state is one in which the atom is at one of its levels (the nth, say) and there are no photons. In the intermediate states, the atom can be at any of its levels, and the field contains one photon. We can say, in

---

† This derivation was first given by H. A. Bethe in 1947, and provided the initial stimulus for the whole subsequent development of quantum electrodynamics.

intuitive terms, that the correction to the energy is due to the emission and subsequent absorption of virtual photons by the atom.[†]

The matrix elements of the electromagnetic interaction operator corresponding to the emission of the photon are, in the non-relativistic case, according to (97.2) and (97.1),

$$-e\sqrt{\frac{2\pi}{\omega\Omega}}(\mathbf{e}^*\cdot\mathbf{v}_{nm}).$$

The summation over intermediate states includes summation over the states of the atom (denoted by the suffix m), integration over the photon momenta (i.e. over $\Omega\, dk_x\, dk_y\, dk_z/(2\pi)^3$), and summation over the photon polarisations. The integration over the directions of **k** and the summation over the polarisations are carried out in the same way as in the derivation of (97.4), and the resulting correction to the energy is

$$-\frac{2e^2}{3\pi}\sum_m\int\frac{|\mathbf{v}_{nm}|^2\,\omega\, d\omega}{(E_m+\omega)-E_n}, \tag{106.1}$$

where E_n and E_m are the unperturbed energy levels of the atom. This integral, however, diverges at the upper limit.

For a free electron, the expression (106.1) would give the correction to the mass, and the renormalisation would consist in simply omitting it, since the "unperturbed" mass of the electron is its observed value. On the other hand, for a free electron the velocity operator $\hat{\mathbf{v}} = \hat{\mathbf{p}}/m$ has only diagonal matrix elements $\mathbf{v}_{nn}$, which coincide with the definite values of $\mathbf{v}$ for a free particle. The sum over m in (106.1) then reduces to the one term $m = n$:

$$-\frac{2e^2}{3\pi}\int\mathbf{v}^2\, d\omega.$$

The renormalisation constant for an electron that is bound (in an atom) is obtained by replacing the squared velocity $\mathbf{v}^2$ by its mean

---

[†] In the non-relativistic theory, the virtualness of the photon is shown by the non-conservation of energy when it is emitted or absorbed. There is no production of virtual electron–positron pairs in the non-relativistic approximation.

value in the relevant state of the atom, i.e. by the matrix element $(\mathbf{v}^2)_{nn}$. From the rule of matrix multiplication, we have

$$(\mathbf{v}^2)_{nn} = \sum_m \mathbf{v}_{nm} \cdot \mathbf{v}_{mn} = \sum_m |\mathbf{v}_{nm}|^2 .$$

Thus we arrive at the expression

$$-\frac{2e^2}{3\pi} \sum_m \int |\mathbf{v}_{nm}|^2 \, d\omega,$$

which must be subtracted from (106.1) in order to obtain the observable correction to the energy level:

$$\delta E_n = \frac{2e^2}{3\pi} \sum_m \int \frac{|\mathbf{v}_{nm}|^2 (E_m - E_n)}{E_m - E_n + \omega} \, d\omega. \qquad (106.2)$$

This integral likewise diverges at the upper limit, but only logarithmically; in a consistent relativistic theory, no divergence would remain, but in the non-relativistic theory a good estimate of δE_n can be found by taking the integration in (106.2) from zero to the electron mass m, bearing in mind that the non-relativistic treatment is valid only for photon frequencies $\omega \ll m$ and that the value of the logarithmic integral is not greatly dependent on the choice of the upper limit if this is large compared with all the differences $E_m - E_n$ between energy levels of the atom.

Lastly, replacing the matrix elements of the electron velocity by those of the dipole moment in accordance with (97.1), we have (in ordinary units)

$$\delta E_n = \frac{2}{3\pi\hbar^3 c^3} \sum_m |\mathbf{d}_{nm}|^2 (E_m - E_n)^3 \log \frac{mc^2}{|E_m - E_n|} . \qquad (106.3)$$

This shift depends on all the electron quantum numbers in the atom (the principal quantum number n, the total angular momentum j, and the orbital angular momentum l). After the correction (106.3) is applied, therefore, the previously degenerate levels with the same n and j but different $l = j \pm \frac{1}{2}$ become distinct.†

---

† For example, at the frequency corresponding to the difference of the levels $E(2s_{1/2})$ and $E(2p_{1/2})$, a numerical calculation from (106.3) gives a value of about 1000 MHz; the result of the exact relativistic calculation is 1050 MHz.

INDEX