NEW AGE

SECOND EDITION

Quantum Mechanics

V.K. Thankappan

NEW AGE INTERNATIONAL PUBLISHERS

QUANTUM MECHANICS

THIS PAGE IS
BLANK

QUANTUM MECHANICS

Second Edition

V.K. Thankappan
Department of Physics
University of Calicut, Kerala
India

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS

New Delhi • Bangalore • Chennai • Cochin • Guwahati • Hyderabad
Jalandhar • Kolkata • Lucknow • Mumbai • Ranchi

Visit us at **www.newagepublishers.com**

Published by New Age International (P) Ltd., Publishers

ISBN (13) : 978-81-224-2500-0

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
4835/24, Ansari Road, Daryaganj, New Delhi - 110002
Visit us at **www.newagepublishers.com**

PREFACE TO THE SECOND EDITION

This second edition differs from the first edition mainly in the addition of a chapter on the Interpretational Problem. Even before the printing of the first edition, there was criticism from some quarters that the account of this problem included in the introductory chapter is too sketchy and brief to be of much use to the students. The new chapter, it is hoped, will remove the shortcoming. In addition to a detailed description of the Copenhagen and the Ensemble Interpretations, this chapter also contains a brief account of the Hidden-Variable Theories (which are by-products of the interpretational problem) and the associated developments like the Neumann's and Bell's theorems. The important role played by the Einstein-Podolsky-Rosen Paradox in defining and delineating the interpretational problem is emphasized. Since the proper time to worry over the interpretational aspect is after mastering the mathematical formalism, the chapter is placed at the end of the book.

Minor additions include the topics of Density Matrix (Chapter 3) and Charge Conjugation (Chapter 10). The new edition thus differs from the old one only in some additions, but no deletions, of material.

It is nearly two years since the revision was completed. Consequently, an account of certain later developments like the Greenbetger-Horne-Zeilinger-Mermin experiment [Mermin N.D. *Physics Today* **36** no 4, p. 38 (1985)] could not be included in Chapter 12. It would, however, be of interest to note that the arguments against the EPR experiment presented in Section 12.4 could be extended to the case of the GHZ-Mermin thought-experiment also. For, the quantum mechanically incorrect assumption that a state vector chosen as the eigenvector of a product of observables is a common eigenvector of the individual (component) observables, is involved in this experiment as well.

Several persons have been kind enough to send their critical comments on the book as well as suggestions for improvement. The author is thankful to all of them, and, in particular, to A.W. Joshi and S. Singh. The author is also thankful to P. Gopalakrishna Nambi for permitting to quote, in Chapter 12, from his Ph.D thesis and to Ravi K. Menon for the use of some material from his Ph.D work in this chapter.

January 1993 V.K. THANKAPPAN

THIS PAGE IS
BLANK

PREFACE TO THE FIRST EDITION

This book is intended to serve as a text book for physics students at the M.Sc. and M. Phil (Pre-Ph.D.) degree levels. It is based, with the exception of Chapter 1, on a course on quantum mechanics and quantum field theory that the author taught for many years, starting with 1967, at Kurukshetra University and later at the University of Calicut. At both the Universities the course is covered over a period of one year (or two semesters) at the final year M.Sc. level. Also at both places, a less formal course, consisting of the developments of the pre-quantum mechanics period (1900-1924) together with some elementary applications of Schrödinger's wave equation, is offered during the first year. A fairly good knowledge of classical mechanics, the special theory of relativity, classical electrodynamics and mathematical physics (courses on these topics are standard at most universities) is needed at various stages of the book. The mathematics of linear vector spaces and of matrices, which play somewhat an all-pervasive role in this book, are included in the book, the former as part of the text (Chapter 2) and the latter as an Appendix.

Topics covered in this book, with a few exceptions, are the ones usually found in a book on quantum mechanics at this level such as the well known books by L. I. Schiff and by A. Messiah. However, the presentation is based on the view that quantum mechanics is a branch of theoretical physics on the same footing as classical mechanics or classical electrodynamics. As a result, neither accounts of the travails of the pioneers of quantum theory in arriving at the various milestones of the theory nor descriptions of the many experiments that helped them along the way, are included (though references to the original papers are given). Instead, the emphasis is on the basic principles, the calculational techniques and the inner consistency and beauty of the theory. Applications to particular problems are taken up only to illustrate a principle or technique under discussion. Also, the Hilbert space formalism, which provides a unified view of the different formulations of nonrelativistic quantum mechanics, is adopted. In particular, Schrödinger's and Heisenberg's formulations appear merely as different representations, analogous respectively to the Hamilton-Jacobi theory and the Hamilton's formalism in classical mechanics. Problems are included with a view to supplementing the text.

From its early days, quantum mechanics has been bedevilled by a controversy among its founders regarding what has come to be known as the *Interpretational Problem*. Judging from the number of papers and books still appearing on this topic, the controversy is far from settled. While this problem does not affect either the mathematical framework of quantum mechanics or its practical applications,

a teacher of quantum mechanics cannot afford to be ignorant of it. It is with a view to giving an awareness of this problem to the teacher of this book that Chapter 1 is included (students are advised to read this chapter only at the end, or at least after Chapter 4). The chapter is divided into two parts: The first part is a discussion of the two main contestants in the arena of interpretation—the *Statistical* (or, *Ensemble*) and the *Copenhagen*. In the second part, the path-integral formalism (which is not considered in any detail in this book) is used to show the connection between the ψ-function of quantum mechanics on the one hand and the Lagrangian function L and the action integral S of classical mechanics on the other. This too has a bearing on the interpretational problem. For, the interpretational problem is, at least partly, due to the proclivity of the Copenhagen school to identify ψ with *the particle* (as indicated by the notion, held by the advocates of this school, that observing a particle at a point leads to a "collapse" of the ψ-function to that point!). But the relationship between S and ψ suggests that, just as S in classical mechanics, ψ in quantum mechanics is a function that characterises the *paths* of the particle and that its appearance in the dynamical equation of motion need be no more mysterious than the appearance of S or L in the classical equations of motion.

The approach adopted in this book as well as its level presumes that the course will be taught by a theoretical physicist. The level might be a little beyond that currently followed in some Universities in this country, especially those with few theorists. However, it is well to remember in this connection that, during the last three decades, quantum theory has grown (in the form of quantum field theory) much beyond the developments of the 1920's. As such, a quantum mechanics course at the graduate level can hardly claim to meet the modern needs of the student if it does not take him or her at least to the threshold of quantum field theory.

In a book of this size, it is difficult to reserve one symbol for one quantity. Care is taken so that the use of the same symbol for different quantities does not lead to any confusion.

This book was written under the University Grants Commission's scheme of preparing University level books. Financial assistance under this scheme is gratefully acknowledged. The author is also thankful to the National Book Trust, India, for subsidising the publication of the book.

Since the book had to be written in the midst of rather heavy teaching assignments and since the assistance of a Fellow could be obtained only for a short period of three months, the completion of the book was inordinately delayed. Further delay in the publication of the book was caused in the process of fulfilling certain formalities.

The author is indebted to Dr. S. Ramamurthy and Dr. K.K. Gupta for a thorough reading of the manuscript and for making many valuable suggestions. He is also thankful to the members of the Physics Department, Calicut University, for their help and cooperation in preparing the typescript.

March 1985 V.K. THANKAPPAN

CONTENTS

CHAPTER 1

INTRODUCTION

Quantum theory, like other physical theories, has two aspects: the mathematical and the conceptual. In the former aspect, it is a consistent and elegant theory and has been enormously successful in explaining and predicting a large number of atomic and subatomic phenomena. But in the latter aspect, which "inquires into the objective world hidden behind the subjective world of sense perceptions"[1], it has been a subject of endless discussions without agreed conclusions[2], provoking one to remark that quantum theory appears to be "so contrary to intuition that the experts themselves still do not agree what to make of it"[3]. In the following section, we give a brief account of the genesis of this conceptual problem, which has defied a satisfactory solution (in the sense of being acceptable to all) in spite of the best efforts of the men who have built one of the most magnificent edifices of human thought. And in Section 1.2 is presented a preview of the salient features of the mathematical aspect of the theory.

1.1 THE CONCEPTUAL ASPECT

In order to understand the root cause of the conceptual problem in quantum mechanics, we have to go back to the formative years of the theory. Quantum theory originated at a time when it appeared that classical physics had at last succeeded in neatly categorising all physical entities into two groups: matter and radiation (or field). Matter was supposed to be composed of 'particles' obeying the laws of Newtonian (classical) mechanics. After the initial controversy as to whether radiation consists of 'corpuscles' or 'waves', Fresnel's work[4] on the phenomenon of diffraction seemed finally to settle the question in favour of the latter. Maxwell's electromagnetic theory provided radiation with a theory as elegant as the Lagrangian-Hamiltonian formulation of Newtonian mechanics.

---

1. Landé, A., *Quantum Mechanics* (Pitman Publishing Corporation, New York 1951), p. 7.
2. See, for example, Landé, A., Born, M. and Biem, W., *Phys. Today*, **21**, No. 8, p. 55 (1968) Ballentine, L.E. et al. *Phys. Today*, **24**, No. 4, p. 36 (1971).
3. Dewitt, B., *Phys. Today*, **23**, No. 9, p. 30 (1970).
4. See, Born, M. and Wolf, E., *Principles of Optics* (Pergamon Press, Oxford 1970), IV Edition pp. xxiii–xxiv.

Particles and Waves in Classical Physics

Now, a particle, according to classical physics, has the following characteristics:

*P*1. Besides having certain invariant attributes such as rest mass, electric charge, etc., it occupies a finite extension of space which cannot, at the same time, be occupied by another particle.

*P*2. It can transfer all, or part, of its momentum and (kinetic) energy 'instantaneously' to another particle in a collision.

*P*3. It has a path, or orbit, characterised by certain constants of motion such as energy and angular momentum, and determined by the principle of least action (Hamilton's principle).

On the other hand, a monochromatic harmonic wave motion is characterised by the following:

*W*1. A frequency ν and a wavelength λ, related to each other by

$$\nu\lambda = (\omega/k) = v, \tag{1.1}$$

where, v is the phase velocity of the wave motion.

*W*2. A real (that is, not complex) function $\psi_{\omega,k}(\mathbf{r}, t) = \phi(\mathbf{k}\cdot\mathbf{r} - \omega t)$, referred to as the wave amplitude or wave function, that satisfies the classical wave equation,

$$\frac{\partial^2\phi}{\partial t^2} = v^2\nabla^2\phi. \tag{1.2}$$

From the linearity (for a given ω) of Eq. (1.2) follows a very important property of wave motions[5]: If $\psi_1, \psi_2, \ldots$ represent probable wave motions, then a linear superposition of these also represents a probable wave motion. Conversely, any wave motion could be looked upon as a superposition of two or more other wave motions. Mathematically,

$$\Psi(\mathbf{r}, t) = \Sigma_i c_i\psi_i(\mathbf{r}, t), \tag{1.3}$$

where the c_i's are (real) constants. Eq. (1.3) embodies the *principle of superposition*, expressed in the preceding statements. It is the basis of the phenomenon of *interference*, believed in classical physics to be an exclusive characteristic of wave motions[6].

Now, experimental and theoretical developments in the domain of microparticles during the early part of this century were such as to render the above concepts of particles and waves untenable. For one thing, it was found, as in the case of electron diffraction (Davisson and Germer 1927)[7], that the principle of super-

---

5. In the following, we will suppress the subscripts ω and k, so that $\psi_{\omega,k}(\mathbf{r}, t)$ is written as $\psi(\mathbf{r}, t)$.

6. Classical wave theory also allows for the superposition of wave motions differing in frequencies (and, thus, in the case of a dispersive medium, in phase velocities). Such a superposition leads to a *wave packet* which, unlike monochromatic wave motions, shares the particle's property (*P*1) of being limited in extension (see Appendix C).

7. The experimental discovery of electron diffraction was preceded by theoretical speculation by Louis de Bröglie (1923) that matter-particles are associated with waves whose wavelength λ is related to the particle-momentum p by $\lambda = h/p$, where h is the universal constant introduced earlier by Max Planck (1900).

position plays an important role in the motion of particles also. For another, radiation was found to share property $P2$ listed above as a characteristic of particles (Photoelectric and Compton Effects)[8]. It was, thus, clear that the classical concepts of particles and waves needed modification. It is the extent and the nature of these modifications that became a subject of controversy.

The Two Interpretations

There have been two basically different schools of thought in this connection. One, led by Albert Einstein and usually referred to as the *Statistical* (or *Ensemble*) *Interpretation* of quantum mechanics[9], maintains that quantum theory deals with statistical properties of an *ensemble* of identical (or, 'similarly-prepared') systems, and not with the motion of an individual system. The principle of superposition is, therefore, not in conflict with properties $P1$ and $P2$, though it is not consistent with $P3$. However, unlike $P1$, $P3$ is not really a *defining property* of particles, but is only a statement of the dynamical law governing particles (in classical mechanics). In place of $P3$, quantum theory provides a law which is applicable only to a statistical ensemble and which, of course, reduces to $P3$ as an approximation when conditions for the validity of classical mechanics are satisfied[10].

The other school, led by Niels Bohr and known as the *Copenhagen Interpretation*, advocates radical departure from classical concepts and not just their modification. According to this school, the laws of quantum mechanics, and in particular the principle of superposition, refer to the motion of individual systems. Such a viewpoint, of course, cannot be reconciled with the classical concept of particles as embodied in $P1$. The concept of 'wave particle duality' is, therefore, introduced according to which there are neither particles nor waves, but only (in classical terminology) particle-like behaviour and wave-like behaviour, one and the same physical entity being capable of both. A more detailed account of this interpretation is given in Chapter 12; the reader is also referred to the book by Jammer[11] and the article by Stapp[12].

---

8. It was in explaining the photoelectric effect that Albert Einstein (1905) reintroduced the concept of light corpuscles, originally due to Isaac Newton, in the form of light quanta which were later named *photons* by G.N. Lewis (1926). Prior to this, Max Planck (1900) had introduced the idea that exchange of energy between matter and radiation could take place only in units of $h\nu$, ν being the frequency of the radiation.
9. For a comparatively recent exposition of the Statistical Interpretation, see, L.E. Ballentine, *Revs Mod. Phys.* **42**, 357 (1970).
10. Thankappan, V.K. and Gopalakrishna Nambi, P. *Found. Phys.* **10**, 217 (1980); Gopalakrishna Nambi, P. *The Interpretational Problem in Quantum Mechanics* (Ph. D Thesis : University of Calicut, 1986), Chapter 5.
11. Jammer, M., *The Conceptual Development of Quantum Mechanics* (McGraw-Hill, New York, 1966), Chapter 7.
12. Stapp, H.P., *Amer. J. Phys.* **40**, 1098 (1972).

The Tossing of Coins

It should be emphasized that the dispute between the two schools is not one that could be settled by experiments. For, experiments in the domain of microparticles invariably involve large number of identical systems, and when applied to large numbers, both the interpretations yield the same result. Besides, even if it were possible to make observations on a single isolated particle, the results could not be taken as a contradiction of the Copenhagen Interpretation[13]. The example of the tossing of coins might serve to illustrate this. The law governing the outcome of tossings of identical coins is contained in the following statement: "The probability for a coin to fall with head up is one half". According to the Statistical Interpretation, this statement means that the ratio of the number of tosses resulting in head up to the total number would be one half if the latter is large enough, the ratio being nearer to the fraction half the larger the number of tosses. In any single toss, either the head will be up or it will be down, irrespective of whether somebody is there to observe this fact or not. However, the application of the law would be meaningless in this case since it is incapable of predicting the outcome of a single toss. This incapability might stem from an ignorance of the factors (parameters) that govern, and the way they influence, the motion of the coin. One cannot, therefore, rule out the possibility of a future theory which is capable of predicting the outcome of a single toss, and from which the above-mentioned statistical law could be deduced (see Chapter 12, Section 5).

The Copenhagen Interpretation, on the other hand, insists that the law is applicable to the case of a single toss, but that it is the statement that the coin falls with either head-up *or* head-down that is meaningless. When no observer is present, one can only say that the coin falls with partially (in this case, half) head-up and partially head-down. If an observation is made, of course, it will be found that the coin is either fully head-up or fully head-down but the act of observation (that is, the interaction between the observer and the coin) is held responsible for changing the coin from a half head-up state to a fully head-up state (or a fully head-down state). Agreement with observation is, thus, achieved, but at a heavy price. For, the coin now is not the classical coin which was capable of falling only with head-up or with head-down but not both ways at the same time. Also, the role of the observer is changed from that of a spectator to an active participant who influences the outcome of an observation. Since the law is presumed to govern the outcome of an individual tossing, it follows that the search for a more fundamental theory is neither warranted nor likely to be fruitful.

A Thought Experiment

At this stage, one might wonder why one has to invent such a complicated scheme of explanation as the Copenhagen Interpretation when the Statistical Interpre-

13. According to the Statistical Interpretation, quantum mechanics does not have anything to say about the outcome of observations on a single particle.

tation is able to account for the observed facts without doing any violence to the classical concept of the coin. Unfortunately, phenomena in the world of microparticles are somewhat more complicated than the tossings of coins. The complication involved is best illustrated through the following thought-experiment. Imagine a fixed screen W with two holes A and B (see Fig. 1.1). In front of this

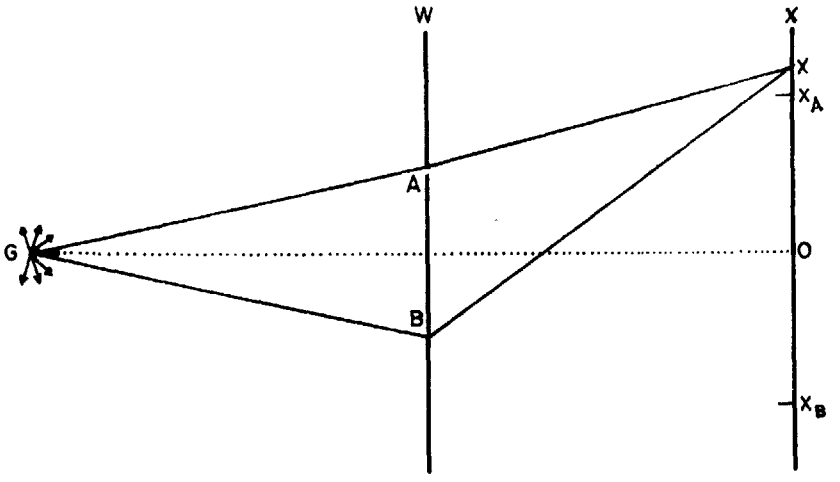


Fig. 1.1. The double slit interference experiment.

screen is an election gun G which shoots out electrons, having the same energy, uniformly in all directions. Behind W is another screen X on which the arrival of the individual electrons can be observed. We first close B and observe the electrons arriving on X for a certain interval of time. We plot the number of electrons versus the point of arrival on X (the screen X will be assumed to be one-dimensional) and obtain, say, the curve I_A shown in Fig. 1.2. Next we close

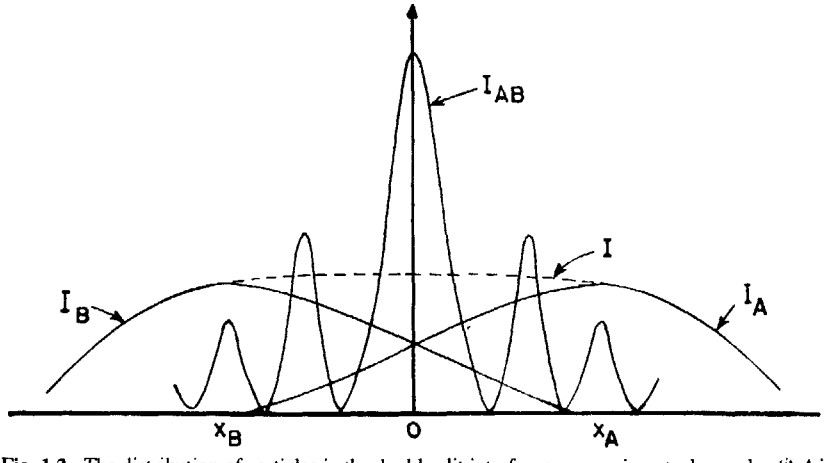


Fig. 1.2. The distribution of particles in the double slit interference experiment when only slit A is open (I_A), when only slit B is open (I_B) and when both A and B are open (I_{AB}). I represents the sum of I_A and I_B.

A and open B and make observation for the same interval of time, obtaining the curve I_B. We now repeat the experiment keeping both A and B open. We should expect to get the curve I which is the sum of I_A and I_B, but get the curve I_{AB} instead. This curve is found to fit the formula

$$I_{AB}(x) = |\psi_A(x) + \psi_B(x)|^2, \tag{1.4}$$

with

$$|\psi_A(x)|^2 = I_A : |\psi_B(x)|^2 = I_B, \tag{1.5}$$

where $\psi_A(x)$ and $\psi_B(x)$ are complex functions of x.

Apparently, our expectation that an electron going through A should not be knowing whether B is closed or open, is not fulfilled. Could it be that every electron speads out like a wave motion after leaving the gun, goes through both the holes and again localises itself on arriving at X? Eqs. (1.4) and (1.5) support such a possibility since these are identical (except for the complex character of ψ_A and ψ_B) with the equations relating amplitudes and intensities of a wave motion. In order to test this, we set up a device near A to observe all the electrons passing through A, before they reach X. We will assume that the electrons arriving on X that are not registered by the device have come through B. We find that the electrons coming through A are, indeed, whole electrons. But, to our surprise, we find that the curves corresponding to the electrons coming through A and B respectively are exactly similar to I_A and I_B, implying that the distribution of electrons on X is now represented not by the curve I_{AB}, but by the curve I. This shows that electrons are particles conforming to the definition $P1$, at least whenever we make an observation on them.

Let us summarise below the main results of the experiment:

$E1$. The number of electrons arriving at a point x on the screen X through A depends on whether B is closed or open. The total number of electrons arriving on X through A is, however, independent of B[14].

$E2$. Observations affect the outcome of experiments.

The results of the electron experiment are easily accommodated in the Copenhagen Interpretation. The basic law governing the electrons in this case is contained in the statement that the probability for an electron that has arrived on X to have come through one of the holes, say A, is P and through the other hole is $(1 - P)$; where $0 \leq P \leq 1$. Since this law governs the motion of each and every electron, when both the holes are open and when no observations are made to see through which hole the electrons are passing, it should be presumed that every electron passes, in a wave-like fashion, through both the holes. Alternatively, one

---

14. This follows from the relation [see Eq. (1.32)],

$$\int_X I_{AB}(x)dx \equiv \int_x |\psi_A(x) + \psi_B(x)|^2 dx$$

$$= \int_X |\psi_A(x)|^2 dx + \int_X |\psi_B(x)|^2 dx$$

$$= \int_X I_A(x)dx + \int_X I_B(x)|dx.$$

could take the view that, as far as the distribution I_{AB} is concerned, the question as to whether a particular electron has come through one or both holes, is not a meaningful one for physics as no experiment can answer the question without affecting the distribution I_{AB}. For any experiment designed to answer the question reveals the electron to be a particle capable of passing through only one hole, but then the distributon is also changed from the one corresponding to classical waves (I_{AB}) to one corresponding to classical particles (I), justifying the hypothesis that the act of observation transforms the electron from a wave-like object extended in space to a particle-like object localised in space. The dichotomy on the part of the electron is easily understood if we realize that particles and waves are merely complementary aspects of one and the same physical entity[15], any one experiment being capable of revealing only one of the aspects and not both[16].

Thus, the Copenhagen Interpretation does not appear so far-fetched when viewed in the context of the peculiar phenomena obtaining in the world of microparticles. However, it denies objective reality to physical phenomena, and prohibits physics from being concerned with happenings in between observations. The question, how is it that the act of observation at one location causes an electron, that is supposed to be spread over an extended space, to shrink to this location?, is dubbed as unphysical. The interpretation, thus, leaves one with an impression that quantum theory is mysterious as no other physical theory is. Those who find it difficult to be at home with this positivist philosophy underlying the Copenhagen Interpretation, will find the Statistical Interpretation more attractive. Let us see how this interpretation copes with the results of the electron experiment.

According to the Statistical Interpretation, the probability law stated earlier as governing the motion of electrons, is a statistical one and is applicable only when a large enough number of 'similarly-prepared' electrons are involved. The distribution of electrons coming through, say hole A, on the screen X being the result of a statistical law, need not be the same when the screen W has only hole A on it as when both A and B are there, just as the distribution of head-up states in the tossings of coins with only one side is different from the distribution of head-up states in the tossings of coins with two sides. Let us elaborate this point: The distribution of electrons coming through hole A on X, is a result of the momentum transfer taking place between the electrons and the screen W at A. The expectation that this momentum transfer, and hence the distribution, are unaffected by the addition of another hole B on W is based on the presumption that a screen with two holes is merely a superposition of two independent screens with one hole each. The experimental result shows that the presumption is not justified. The

5. The Principle of Complementarity, which seeks to harmonize the mutually exclusive notions of particles and waves, was proposed by Neils Bohr (1928). A detailed account of the principle is given in the reference quoted in footnote 11 as well as chapter 12.

6. This limitation on the part of experiments is enshrined in the Uncertainty Principle proposed by Werner Heisenberg (1927), which puts a limit on the precision with which complementary variables such as position and momentum of a particle can be measured.

fact that the momentum transfer at the hole A when both A and B are open is different from the momentum transfer when only A is open, could also be understood on the basis of the quantization of the momentum transfer resulting from the periodicity in space of the holes in the former case (W. Duane 1923)[17].

Thus, experimental result $E1$ is easily understood on the basis of the Statistical Interpretation. As for $E2$, one should distinguish between the two ways in which observations affect the outcome of experiments. One is that observations on electrons coming through hole A affect their distribution on the screen X. This could be understood as due to the fact that the momentum transfer involved in the act of observation is not negligible compared with the momentum of the electrons themselves. The other is that observations on electrons coming through hole A affect (apparently) also the distribution of electrons coming through hole B. In order to accommodate this fact within the framework of the Statistical Interpretation, one has to assume that the statistical correlation that exists between two *paths* (of the electrons), one passing through A and the other through B, is such that it can be destroyed by disturbing only one of the paths. In fact, a correlation represented by the linear superposition of two functions ψ_A and ψ_B as in Eq. (1.3), whose phases are proportional to the classical actions associated with the paths, satisfies such a condition[10]. For, as is known from the classical theory of waves, the correlation can be destroyed by introducing a random fluctuation in the phase of one of the functions. So in order to understand the experimental result, one has to assume that observations on the electrons always introduce such a random variation in the action associated with the path of the electrons[18].

The 'Mystery' in Quantum Mechanics

Thus, in the course of understanding $E2$, we are led to introducing a (complex) function which, in certain aspects such as the applicability of the principle of superposition, resembles a wave amplitude[19]. This is the really new element in quantum mechanics; it represents an aspect of microworld phenomena quite foreign to classical statistical processes such as the tossings of coins. But whereas the Copenhagen school regards these functions as incompatible with the classical

---

17. The period would be the distance d between the holes. According to Duane's hypothesis the momentum transfer between the screen W and the electron, when both A and B are open, has to be an integral multiple of (h/d), h being the Planck's constant. This relationship is identical with the de Bröglie relation, $p = h/\lambda$ (see footnote 7) if we recognise the wavelength λ as a periodicity in space. Duane's hypothesis is an extension, to the case of the linear momentum, of the earlier hypotheses of Max Planck (footnote 8) and of Neils Bohr (1913) on the relationship between the quantization of energy and periodicity τ in time [energy = integral multiple of (h/τ)] and quantization of angular momentum and periodicity 2π in angles [angular momentum = integral multiple of $(h/2\pi)$], respectively.
18. This is nothing but the Uncertainty Principle.
19. Erwin Schrödinger (1926) was the first to introduce these functions and to derive an equation of motion (the *Schrödinger equation*) for them. The physical interpretation of these functions as probability amplitudes which are related to the probability of finding the particles at a space point in the same way as wave amplitudes are related to wave intensities, is due to Max Born (1926).

concept of particles, and invests them with a certain amount of physical reality, thereby endowing quantum mechanics with an aura of mystery, the Statistical Interpretation makes a distinction between these functions and the physical entities involved. The physical entities are the electrons or other microparticles (conforming to definition *P*1), but the functions are mathematical entities characterising the paths of the microparticles just as the action in classical mechanics is a mathematical function characterising the classical paths of particles. The functions, thus, determine the dynamical law governing the motion of microparticles. This law is, admittedly, new and different from the dynamical law in classical mechanics. But, then, it is not the first time in physics that a set of rules (theory) found to be adequate for a time, proved to be inadequate in the light of new and more accurate experimental facts. Also, the fact that quantum mechanics does not provide an *explanation* to the dynamical law or laws (such as the principle of superposition) underlying it, does not justify alleging any special mystery on its part, since such mysteries are parts of every physical theory. For example, classical mechanics does not explain why the path of a particle is governed by Hamilton's principle, eletromagnetic theory does not offer an explanation for Coulomb's or Faraday's laws and the theory of relativity does not say why the velocity of light in vacuum is the same in all inertial frames. Thus, from the viewpoint of the Statistical Interpretation, quantum mechanics is no more mysterious than other physical theories are. It certainly represents an improvement over classical mechanics since it is able to explain Hamilton's principle, but an explanation of the fundamental laws underlying quantum mechanics themselves need be expected only in a theory which is more fundamental than quantum mechanics.

It should be clear from the foregoing discussion that the choice between the Copenhagen and the Statistical Interpretations could be one of individual taste only. Anyway, the mathematical formalism of quantum mechanics is independent of these interpretations.

1.2 THE MATHEMATICAL ASPECT

One or the other branch of mathematics plays a dominant role in the formulation of every physical theory. Thus, classical mechanics and electromagnetic theory rely heavily on differential and vector calculus, while tensors play a dominant role in the formulation of the general theory of relativity. In the case of quantum mechanics, it is the mathematics of the infinite-dimensional linear vector spaces (the Hilbert space) that play this role. In this section, we will show how the basic laws of quantum mechanics[20] make this branch of mathematics the most appropriate language for the formulation of quantum mechanics.

---

20. In the form originally proposed by Feynman, R.P. [*Revs. Mod. Phys.* **20**, 367 (1948); also, Feynman, R.P. and Hibbs, A.R., *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York 1965)] and later modified by V.K. Thankappan and P. Gopalakrishna Nambi[10]. The basic laws of non-relativistic quantum mechanics were discovered during the period 1900-1924 through the efforts of many physicists, and a consistent theory incorporating these laws were formulated during the period 1925-1926 mainly by Erwin Schrödinger (1926) in the form of *Wave Mechanics* and by Werner Heisenberg, Max Born and Pascal Jordan (1925–1926) in the form of *Matrix Mechanics*.

Now, in classical mechanics the motion of a particle is governed by the Principle of Least Action (Hamilton's Principle). According to this principle, the path of a particle between two locations A and Q in space is such that the action $S(Q, t_Q : A, t_A)$ defined by,

$$S(Q, t_Q : A, t_A) = \int_{t_A}^{t_Q} L\,dt = \int_A^Q p\,dq - \int_{t_A}^{t_Q} H\,dt, \tag{1.6}$$

is a minimum, where L is the Lagrangian, p the momentum and H the Hamiltonian of the particle, and t_A and t_Q are, respectively, the time of departure from A and the time of arrival at Q. Thus, the path between A and Q is determined by the variational equation,

$$\delta S = 0. \tag{1.7}$$

We will call the path defined by Eq. (1.7) the *classical path* and will denote it by α_c and the action corresponding to it by $S_c(Q, t_Q : A, t_A)$.

As we have already mentioned, experiments in the domain of microparticles have shown that the paths of these particles are not governed by the principle of least action. However, the results of these experiments are consistent with, indeed suggestive of, the following postulates which could be regarded as the quantum mechanical laws of motion applicable to microparticles:

*Q*1. Associated with every path α of a particle[21] from location A to location Q in space, is a complex function $\phi_\alpha(Q, t_Q : A, t_\alpha)$ given by,

$$\phi_\alpha = a_\alpha \exp\,[(i/\hbar)S_\alpha], \tag{1.8}$$

where

$$S_\alpha(Q, t_Q : A, t_\alpha) = \int_{t_\alpha}^{t_Q} L\,dt = \int_A^Q p\,dq - \int_{t_\alpha}^{t_Q} H\,dt. \tag{1.9}$$

t_α, here, has the same meaning as t_A in Eq. (1.6) except that it could be different for the different paths α. Also[22], $\mathrm{h} = h/2\pi$.

*Q*2. The *probability amplitude* for a particle to go from A (at some time) to Q at time t_Q is $\psi_A(Q, t_Q)$, where,

$$\psi_A(Q, t_Q) = \Sigma_\alpha \phi_\alpha(Q, t_Q : A, t_\alpha). \tag{1.10}$$

*Q*2a. Only those paths contribute to the summation in Eq. (1.10) that differ from α_c by less than $\hbar/2$ in action. That is

$$\Delta S_\alpha \equiv (S_\alpha - S_c) < (\hbar/2). \tag{1.10a}$$

*Q*3. If $A, B, C, \ldots$ are locations corresponding to similarly prepared states[23] of a particle in an experimental set up, the *number* of particles arriving at a point of a observation, Q, at time t_Q from the above locations, is proportional to $|\Psi(Q, t_Q)|^2$, where,

$$\Psi(Q, t_Q) = \Sigma_A c_A \Psi_A(Q, t_Q), \tag{1.11}$$

---

21. We assume that the spin of the particle is zero.
22. The one-letter notation for $(h/2\pi)$ was first introduced by P.A.M. Dirac (1926), in the form "h". For this reason, $\hbar$ is also called Dirac's constant.
23. This phrase stands for 'elements, or members, of an ensemble'.

the c_A's being numbers (in general, complex) to be chosen such that

$$\int |\Psi|^2 d^3\mathbf{r}_Q = 1, \tag{1.11a}$$

where $d^3\mathbf{r}_Q$ represents an element of volume containing the point Q.

If α is a path between A and Q, and β a path between B and Q, then, as a consequence of condition (1.10a), we will have,

$$|(S_\alpha - S_\beta)| - \Delta S_{AB} < (\hbar/2), \tag{1.12a}$$

where

$$\Delta S_{AB} = |S_{\alpha_c} - S_{\beta_c}|.$$

α_c and β_c being the classical paths between A and Q and between B and Q, respectively. Also, corresponding to every path 'a' between A and Q (that contributes to ψ_A), there will be a path 'b' between B and Q such that

$$|S_\alpha - S_b| = \Delta S_{AB}. \tag{1.12b}$$

Eq. (1.12b) enables us to say that the phase difference between ψ_A and ψ_B is the quantity $(\Delta S_{AB}/\hbar)$ whereas inequality (1.10a), from which inequality (1.12a) follows, ensures that the phase difference is such a definite quantity. Now, a definite phase difference between ψ_A and ψ_B is the condition for A and B to be *coherent sources* (or, similarly-prepared states) from the viewpoint of Q. We will, therefore, refer to inequality (1.10a) as the *coherency condition.*

Postulate $Q3$ incorporates the principle of superposition referred to in Section 1.1 (Eq. (1.3)). However, unlike c_i and ψ_i in (1.3), c_A and ψ_A in Eq. (1.11) are complex quantities. Therefore, it is not possible to interpret ψ_A and Ψ in (1.11) as representing wave motions in the physical space[24]. Also, the principle of superposition will conflict with property $P1$ of particles (see, p. 2), if applied to the case of a single particle. But there is no experimental basis for invalidating $P1$; on the contrary, experiments confirm the continued validity of $P1$ by verifying, for example, that all electrons have the same spin, (rest) mass and electric charge both before and after being scattered by, say, a crystal. Therefore, the principle of superposition should be interpreted as applying to the statistical behaviour of a large number (ensemble) of identical systems. In fact, the terms 'probability amplitude' and 'number of particles' emphasize this statistical character of the postulates. However, the really new element in the theory is not its statistical character, but the law for combining probabilities. Whereas in the classical statistics, probabilities for independent events are added to obtain the probability for the combined event (If $P_A(Q)$ and $P_B(Q)$ are, respectively, the probabilities for the arrival of a particle at Q from A and from B, then, the probability $P_{AB}(Q)$ for the arrival of a particle at Q from either A or B is given by $P_{AB}(Q) = P_A(Q) + P_B(Q)$) in the new theory, this is not always so. In particular, whenever criterion (1.12a)

---

24. In classical wave theory also, complex amplitudes are employed sometimes, purely for the sake of calculational convenience. Care is, then, taken in the computation of physically significant quantities such as the intensity of the wave motion, to separate out the contribution due to the imaginary part of the amplitudes. In quantum mechanics the ψ_A's are perforce complex. [see Eq. (4.15 b)]

is satisfied, it is the probability *amplitudes* that are to be added in place of the probabilities [If $\psi_A(Q)$ and $\psi_B(Q)$ are the probability amplitudes for the arrival of a particle at Q from A and from B, respectively, then the probability amplitude for the arrival of a particle at Q from either A or B is given by $\psi_{AB}(Q) = \psi_A + \psi_B$, so that $P_{AB}(Q) = |\psi_{AB}(Q)|^2 = P_A(Q) + P_B(Q) + 2\,\mathrm{Re}(\psi_A^* \psi_B) \neq P_A + P_B$]. It appears that events which are supposed to be independent, are not really so.

If it appears odd that the dynamics of particles should be governed by such abstractly-defined entities as the probability amplitudes, one has only to observe that these functions are not any more odd than the classical action in terms of which they are defined and which plays an important role in the dynamics of classical particles. It also turns out that the differential equations satisfied by these probability amplitudes are closely related to the Hamilton-Jacobi equations satisfied by the action in classical mechanics, as is being demonstrated below:

The Schrödinger Equation

In view of (1.10a), the contribution to ψ_A in (1.10) from paths lying in a single plane may be written as,

$$\psi_A'(\mathbf{r}_Q, t_Q) = 2\Big[\int_{S_c}^{S_c+\hbar/2} a(S)\rho(S) \exp\{(i/\hbar)S\}\, dS - a(S_c+\hbar/2) \exp\{(i/\hbar)(S_c+\hbar/2)\}\Big] - a(S_c) \exp\{(i/\hbar)S_c\}, \tag{1.13a}$$

where the last term compensates for counting the path α_c twice in the integral. Also, $\rho\,(S)\,dS$ represents the number of paths in the plane having action between S and $S + dS$, while $a(S_\alpha) = a_\alpha$. We make the assumption that both $a(S)$ and $\rho(S)$ are independent of S. This means that : (i) all the paths between A and Q that satisfy condition (1.10a) contribute equally to ψ_A, (ii) the density of paths is uniform in the S-space. Eq. (1.13a) then reduces to,

$$\psi_A'(\mathbf{r}_Q, t_Q) = A' \exp\{(i/\hbar)S_c(\mathbf{r}_Q, t_Q)\}, \tag{1.13b}$$

A' being a (complex) constant.

The total contribution to ψ_A could, in principle, be written as the sum of the contributions from the various planes. Under the above assumptions, the contribution from each plane would be proportional to $\exp[(i/\hbar)S_c]$, so that

$$\psi_A(\mathbf{r}_Q, t_Q) = A \exp\{(i/\hbar)S_c(\mathbf{r}_Q, t_Q)\}, \tag{1.13c}$$

In Eqs. (1.13a-c), $\mathbf{r}_Q$ is the position vector of the location Q.

For a fixed location A and variable locations Q, Eq. (1.13c) yields,

$$i\hbar \frac{\partial \psi_A(\mathbf{r}_Q, t_Q)}{\partial t_Q} = H\psi_A(\mathbf{r}_Q, t_Q), \tag{1.14a}$$

$$(\hbar/i)\nabla_Q \psi_A(\mathbf{r}_Q, t_Q) = \mathbf{p}\psi_A(\mathbf{r}_Q, t_q), \tag{1.14b}$$

where the Hamilton-Jacobi equations[25],

$$\frac{\partial S_c}{\partial t_Q} = -H(Q), \tag{1.15a}$$

$$\nabla_Q S_c = \mathbf{p}(Q), \tag{1.15b}$$

have been used. Here, H and $\mathbf{p}$ are, respectively, the Hamiltonian and the linear momentum (along α_c) of the particle at Q. Since the location A is arbitrary we have, from Eqs. (1.14a) and (1.11),

$$i\hbar \frac{\partial \Psi(\mathbf{r}_Q, t_Q)}{\partial t_Q} = H\Psi(\mathbf{r}_Q, t_Q). \tag{1.16}$$

Erwin Schrödinger was the first to derive this equation (see footnote no. 19), on the basis of de Bröglie's theory of matter waves and Hamilton's analogy of his principle to Fermat's principle of at least time in optics[26]. It is, therefore, known as *Schrödinger's Wave Equation,* while Ψ is called Schrödinger's *wave function.* The equation describes the evolution of Ψ in time and is, thus, the *equation of motion* for Ψ.

The Uncertainty Principle

Inequalities (1.10a) and (1.12a) could be interpreted as implying that the *random fluctuations* ΔS_A in the actions associated with each of the amplitudes ψ_A should be less than $(\hbar/2)$ for them to be superposable as in Eq. (1.11). Conversely, if

$$\Delta S_A \geq (\hbar/2), \tag{1.17}$$

then, ψ_A cannot be superposed with other amplitudes (that is, there is no 'interference' between ψ_A and other amplitudes). Experimentally, it is found that a successful attempt at observing the paths associated with ψ_A invariably leads to the destruction of the interference between ψ_A and other amplitudes. This means that *two interfering 'paths' cannot be observed without destroying the interference between them.* Werner Heisenberg (see, footnote no. 16) was the first to recognise this fact[27] and to suggest a mathematical expression for it, in the form (see, Ref. 11, Section 7.1),

$$\Delta q \cdot \Delta p \geq h, \tag{1.18}$$

---

25. See Landau, L.D. and Lifshitz, E.M., *Mechanics* (Pergamon Press, (1969)), II Edition, Sections 43 and 47.
26. See Ref. 11, Section 5.3.
27. Unfortunately, there are different interpretations for Heisenberg's Uncertainty Principle [see H. Margenau and L. Cohen, *Probabilities in Quantum Mechanics,* in **Quantum Theory and Reality,** (Ed) M. Bunge (Springer-Verlag, 1967), chapter 4], just as there are different interpretations of quantum mechanics itself. The version given here is the one that follows naturally from the postulates.

where Δq and Δp are the uncertainties [hence the name, 'Uncertainty Principle' for the expression (1.18)] in the coordinate q and the momentum p conjugate to q, respectively. Now, the uncertainties are defined as the root-mean-square (or, standard) deviations from the average values. With this definition, it is possible to reduce (see Ref. 10) the semi-equality (1.17) to an expression similar to (1.18), namely,

$$\Delta q_A \, \Delta p_A \geq \hbar/2, \tag{1.17a}$$

where Δq_A and Δp_A are, respectively, the root-mean-square values of the displacements and the momenta (of the paths associated with ψ_A) perpendicular to the classical path (the 'average' path) belonging to ψ_A. The uncertainty principle, thus, has its root in the coherency condition.

The Meaning of Ψ

According to postulate $Q3$, the number of particles coming, at time t_Q, into a volume element $dV = d^3\mathbf{r}_Q$ containing the location Q, from the various sources A, B, C, ..., is proportional to $|\Psi(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q$. Therefore, the total number N of particles in the volume V is given by

$$N = a||\Psi||^2, \tag{1.19}$$

where 'a' is a proportionality constant, and

$$||\Psi||^2 = \mathrm{Lt}_{V\to\infty} \int_V |\Psi(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q. \tag{1.20}$$

In practice, V need not be infinite, but should be large enough for Ψ to vanish outside V.

It is customary to define Ψ such that $a = N$ in Eq. (1.19). Then, we have,

$$||\Psi||^2 = 1. \tag{1.21}$$

The interpretation of this equation is that $||\Psi||^2$ is the *probability* of finding a particle in the volume V. The fact that Ψ is not zero implies that there is a particle in the volume V, so that the particle should be *certainly* found somewhere in V. $|\Psi(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q$ then is the probability of finding the particle in the volume element $d^3\mathbf{r}_Q$, and $|\Psi(\mathbf{r}_Q, t_Q)|^2$ is the *probability density*[28]. That is, $N|\Psi(\mathbf{r}_Q, t_Q)|^2$ is the number of particles in a unit volume containing the location Q.

A similar interpretation could be given to the ψ_A's : $|\psi_A(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q$ is the probability that a particle, whose source is A, is found in the volume element $d^3\mathbf{r}_Q$ at time t_Q and $\int|\psi_A(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q$ is the probability that the particle is found somewhere in the volume V. Thus,

---

28. The interpretation of $|\Psi|^2$ as a probability density is originally due to Max Born (see footnote 19).

$$||\psi_A||^2 \equiv \mathrm{Lt}_{V\to\infty} \int_V |\psi_A(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q = 1, \tag{1.22}$$

and

$$N_A = N_A ||\psi_A||^2, \tag{1.23}$$

where N_A is the number of particles in the volume V, having their source at A.

In applying the foregoing interpretation of ψ_A, however, we have to be careful not to contradict experiments. For, experiments show that the *distribution* of the particles in V, originating from A, is affected by the presence of other sources satisfying condition (1.12a). In fact (see Fig. 1.2),

$$N|\Psi(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q \neq \Sigma_A N_A |\psi_A(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q. \tag{1.24}$$

Therefore, $|\psi_A(\mathbf{r}_Q, t_Q)|^2 d^3\mathbf{r}_Q$ is the probability for the particle from A to be found at time t_Q in the volume element $d^3\mathbf{r}_Q$ only when there is no other coherent sources present.

A modification in the distribution of the particles from A does not, however, affect the total number of particles in V that have their source at A. Thus, Eqs. (1.22) and (1.23) are valid whether there are other sources or not. Therefore, the total number of particles in V can be written as,

$$N = \Sigma_A N_A, \tag{1.25}$$

or,
$$N||\Psi||^2 = \Sigma_A N_A ||\psi_A||^2,$$

so that
$$||\Psi||^2 = \Sigma_A (N_A/N) ||\psi_A||^2. \tag{1.26}$$

But, from Eq. (1.11), we have,

$$||\Psi||^2 = \Sigma_A \Sigma_B c_A^* c_B \int_V \psi_A^*(\mathbf{r}_Q, t_Q)\psi_B(\mathbf{r}_Q, t_Q) d^3\mathbf{r}_Q. \tag{1.27}$$

Comparing Eqs. (1.26) and (1.27), we get[29],

$$|c_A|^2 = (N_A/N), \tag{1.28}$$

$$\int_V \psi_A^*(\mathbf{r}_Q, t_Q)\psi_B(\mathbf{r}_Q, t_Q)\, d^3\mathbf{r}_Q = \delta_{AB}. \tag{1.29}$$

where δ_{AB} is the Kronecker delta function :

$$\delta_{AB} = 0, \text{ if } A \neq B,$$

$$= 1, \text{ if } A = B. \tag{1.30}$$

From Eqs. (1.28) and (1.25), we have

$$\Sigma_A |c_A|^2 = 1, \tag{1.31}$$

---

29. In place of (1.29), we can have the condition

$$c_A^* c_B \int \psi_A^* \psi_B d^3\mathbf{r}_Q + c_B^* c_A \int \psi_B^* \psi_A d^3\mathbf{r}_Q = 0,$$

which is equivalent to requiring that the real part of

$$\int (c_A \psi_A)^* (c_B \psi_B) d^3\mathbf{r}_Q \text{ be zero}.$$

while, from Eqs. (1.27) and (1.29), we see that Eq. (1.31) is just another expression for Eq. (1.21).

The Algebra Obeyed by ψ

Now, Eq. (1.27) with condition (1.29) can be expressed as,

$$||\Sigma_A c_A \psi_A||^2 = \Sigma_A ||c_A \psi_A||^2. \tag{1.32}$$

Actually, Eq. (1.29) does not fully express the conditions on the integral $\int \psi_A^* \psi_B d^3\mathbf{r}_Q$ that are implicit in the relation (1.32). These conditions are listed below, where we use the abbreviation,

$$(\psi_A, \psi_B) \equiv \int \psi_A^*(\mathbf{r}_Q, t_Q)\psi_B(\mathbf{r}_Q, t_Q)d^3\mathbf{r}_Q, \tag{1.33}$$

$$(\psi_A, \psi_B) = (\psi_B, \psi_A)^*, \tag{1.34^1}$$

$$(\psi_A, c_B\psi_B) = c_B(\psi_A, \psi_B), \tag{1.34^2}$$

$$(\psi_A, \psi_B + \psi_C) = (\psi_A, \psi_B) + (\psi_A, \psi_C), \tag{1.34^3}$$

$$||\psi_A||^2 \geq 0, \tag{1.34^4}$$

the equality sign in (1.34^4) holding only when ψ_A is a null function (that is, when $\psi_A \equiv 0$). The restriction of $||\psi_A||^2$ to positive values is essential for its probabilistic interpretation.

From Eqs. (1.29) and (1.11), we have

$$c_A = (\psi_A, \Psi) \tag{1.35}$$

so that, Eq. (1.31) becomes,

$$\Sigma_A(\Psi, \psi_A)(\psi_A, \Psi) = 1 \tag{1.36}$$

The (scalar) number[30] (1.34^1) is called the *scalar* or *inner-product* of ψ_A and ψ_B. This terminology anticipates the possibility of regarding the function ψ_A as a vector in some function space. In fact, such a possibility is strongly suggested by the following comparison.

Let X be a vector in the (3-dimensional) physical space. Then,

$$\mathbf{X} = \sum_{k=1}^{3} x_k \mathbf{e}_k, \tag{1.37}$$

where $\mathbf{e}_k$ (k = 1, 2, 3) are the unit vectors along three mutually perpendicular directions, so that,

$$(\mathbf{e}_k, \mathbf{e}_j) = (\mathbf{e}_j \cdot \mathbf{e}_k) = \delta_{jk}, \tag{1.38}$$

---

30. It is possible to have functions for which (1.29) is not the Kronecker delta but some other scalar number (see Eq. (1.35) and footnote 29), but which otherwise satisfy all the conditions (1.34$^{1\text{-}4}$).

$(\mathbf{e}_j \cdot \mathbf{e}_k)$ being the scalar product of $\mathbf{e}_j$ and $\mathbf{e}_k$. x_k is the component of $\mathbf{X}$ along $\mathbf{e}_k$. Obviously,

$$x_k = (\mathbf{e}_k \cdot \mathbf{X}). \tag{1.39}$$

If $\mathbf{X}$ is a unit vector, then

$$\sum_{k=1}^{3} x_k^2 \equiv \Sigma_k (\mathbf{X} \cdot \mathbf{e}_k)(\mathbf{e}_k \cdot \mathbf{X}) = 1 \tag{1.40}$$

The similarity of Eqs. (1.11), (1.29), (1.35) and (1.36), respectively, to Eqs. (1.37), (1.38), (1.39) and (1.40) is obvious, and suggests that Ψ could be regarded as a vector with 'components' c_A along the directions of the 'unit vectors' ψ_A. Since, however, c_A and ψ_A, unlike x_k and $\mathbf{e}_k$, are complex quantities and since the number of independent ψ_A's [that is, those satisfying condition (1.29)] is not limited to three, Ψ is a vector, not in the physical space, but in some abstract function space. The nature of this space could be inferred from the properties of the ψ_A's. In enumerating these properties, we will adopt the following definitions and notations:

$\{\psi_A\}$ will denote the set of independent functions ψ_A that occur in the expansion of Ψ as in Eq. (1.11). Ψ will stand for any function whose scalar products with two or more members of the set ψ_A, are non-zero. The ψ_A's and the Ψ's together form a family of functions denoted by $[\Psi, \psi_A]$.

The properties of the ψ_A's and the Ψ's could be now summarised as follows:

*V*1. The sum of two or more members of the family is a Ψ and, thus, belongs to the family. Obviously,

$$\psi_A + \psi_B = \psi_B + \psi_A,$$

and $\psi_A + (\psi_B + \psi_C) = (\psi_A + \psi_B) + \psi_C = \psi_A + \psi_B + \psi_C$.

*V*2. It is possible to have a function Ψ which is identically zero. For example, in the case of the two-slit interference experiment [see Eq. (1.4)], we do have, $I_{AB}(x) = |\psi_A(x) + \psi_B(x)|^2 \equiv |\Psi|^2 = 0$. Such functions will be called *null* and will be denoted by **0**.

*V*3. According to the postulates, only $|\psi_A|^2$ and $|\Psi|^2$ have direct physical meaning (being related to the number of particles). Therefore it is possible to associate a $-\psi_A$ with every $+\psi_A$ and a $-\Psi$ with every $+\Psi$ such that $+\psi_A - \psi_A = \mathbf{0}$ and $+\Psi - \Psi = \mathbf{0}$. Thus with every function in $[\Psi, \psi_A]$, we can associate an *additive inverse*.

*V*4. Multiplication of Ψ by a scalar c yields a function $\Psi = c\Psi$ which differs from Ψ only in that the proportionality constant relating $||\Psi||^2$ to the number of particles is different from that appearing in Eq. (1.19). But Eq (1.16) shows that this proportionality constant has no effect on the dynamics of the system. Therefore, $c\Psi$ belongs to $[\Psi, \psi_A]$. Similar remark apply to $c\psi_A$.

The above properties of ψ_A and Ψ show that, if these functions are to be regarded as vectors in some abstract space, then, that space should be *closed* under vector-addition and under multiplication by scalars, that there should be a *null-vector* in the space and that every vector in the space should be associated with an *additive inverse*. These are the properties of a *linear vector space* (see, Section 2.1). Furthermore, we have seen that it is possible to define a *scalar product* in the space with properties (1.34[1-4]). Therefore, the space is *unitary*.

If Ψ and ψ_A (and hence ϕ_α) are to be regarded as vectors, then, Eqs. (1.16) and (1.14b) show that H and $\mathbf{p}$ should be identified with the differential *operators* $i\hbar \frac{\partial}{\partial t}$ and $(\hbar/i)\nabla$, respectively. Since H and $\mathbf{p}$ are (in classical mechanics) dynamical variables of a mechanical system, it is reasonable to expect that other dynamical variables, such as angular momentum and position co-ordinates, are also represented by operators (not, necessarily, differential) that act on the space pertaining to the system. It is, in fact, found possible to develop an elegant and powerful formalism of quantum mechanics based on the above concepts of Ψ as a vector and the dynamical variables as operators in an infinite-dimensional, unitary, linear vector space (usually referred to as the *Hilbert* space). We will devote the next three chapters to such a formulation of the basic principles of quantum mechanics, an outline of which has been presented in this chapter.

GLOSSARY OF ORIGINAL PAPERS

Bohr, Neils. 1913. *Philosophical Magazine*, **26**, 1; 476; 857. See also, *On the Constitution of Atoms and Molecules* (W.A. Benjamin, New York, 1963).
Bohr, Neils. 1928. *Nature*, **121**, 580. Also *Die Naturwissenschaften*, **17**, 483 (1929); *Atomic Theory and the Description of Nature* (Cambridge University Press, 1961).
Born, Max. 1926. *Zeitschrift für Physik*, **37**, 863; **38**, 803.
Born, M., Heisenberg, W. and Jordan, P. 1926. *Zeitschrift für Physik*, **35**, 557.
Born, Max and Jordan, Pascal. 1925. *Zeitschrift für Physik*, **34**, 858.
Davisson, C.J. and Germer, L.H. 1927. *Physical Review*, **30**, 705.
De Bröglie, Louis. 1923. *Comptes Rendus*, **177**, 507; 548; 630.
Dirac, P.A.M. 1926. *Proceedings of the Royal Society of London* (A) **110**, 561.
Duane, W. 1923. *Proceedings of the National Academy of Sciences*, **9**, 158.
Einstein, Albert. 1905. *Annalen der Physik*, **17**, 132. (see also, A.B. Arons and M.B. Peppard, *American Journal of Physics*, **33**, 367 (1965)).
Heisenberg, Werner. 1925. *Zeitschrift für Physik*, **33**, 879.
Heisenberg, Werner. 1927. *Zeitschrift für Physik*, **43**, 172.
Lewiss, G.N. 1926. *Nature*, **118**, 874.
Planck, Max. 1900. *Verhandlungen der Deutschen Physikalischen Gesselschaft*, **2**, 237; Also, *Annalen der Physik*, **4**, 553 (1901).
Schrödinger, Erwin. 1926. *Annalen der Physik*, **79**, 361; 489; **80**, 437; **81**, 109. The time-dependent Schrödinger equation was given in the last one of these papers.

CHAPTER 2

LINEAR VECTOR SPACES

We have seen, in the previous chapter, that the probability amplitudes $\psi(\mathbf{r}, t$ could be regarded as vectors in a linear vector space. In this chapter, we will develop the mathematical formalism underlying linear vector spaces.

2.1 VECTORS

A linear vector space $\mathcal{V}$ is defined by a *set*[1] S of element $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$ called *vectors* and a field $\mathcal{F}$ of numbers $a, b, c, \ldots$ called *scalars*, with the following properties:

(i) The space is *closed* under vector-addition. That is, corresponding to any two vectors $\mathbf{x}$ and $\mathbf{y}$ in the space, there is a *unique* third vector $\mathbf{z}$ which is also in the space, such that

$$\mathbf{x} + \mathbf{y} = \mathbf{z}. \tag{2.1}$$

Vector-addition is *commutative*:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}, \tag{2.2}$$

and *associative*:

$$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + \mathbf{y} + \mathbf{z}; \tag{2.3}$$

for any three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$.

(ii) There is a *null* or *zero vector* $\mathbf{0}$, such that, for any vector $\mathbf{x}$ in the space,

$$\mathbf{x} + \mathbf{0} = \mathbf{x}. \tag{2.4}$$

The null vector defines the *origin* of the vector space.

(iii) Corresponding to every $\mathbf{x}$, there is an additive *inverse* $(-\mathbf{x})$, such that,

$$\mathbf{x} + (-\mathbf{x}) = \mathbf{0}. \tag{2.5}$$

---

1. A set $S \equiv \{s_i\}$ is a collection of objects s_i $(i = 1, 2, \ldots)$, which are called *elements* of the set, connected by some common attribute. Examples are a set of real numbers, a set of atoms, etc. The number n of elements in S is called the *cardinal number* of S. The set is finite or infinite depending on whether n is finite or infinite. S_i is a *subset* of S if every element in S_i is also an element of S but not vice versa. Thus, the set of positive integers is a subset of the set of integers. An infinite set has at least one subset which has got the same cardinal number as the original set. For example, the set of perfect squares is a subset of the set of positive integers. The cardinal number is the same for both the sets since there is a perfect square corresponding to every positive integer.
A set is *denumerable* (countable) if it has the same cardinal number as the set of positive integers; otherwise it is *nondenumerable*.

(iv) The space is closed under multiplication by scalars. That is, for any scalar c, the vector $\mathbf{y} = c\mathbf{x}$ is in the space if $\mathbf{x}$ is in the space.
Multiplication by scalars is commutative:

$$c\mathbf{x} = \mathbf{x}c, \tag{2.6}$$

associative:

$$a(b\mathbf{x}) = (ab)\mathbf{x} = ab\mathbf{x}, \tag{2.7}$$

and *distributive*:

$$(a+b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}; \tag{2.8a}$$

$$a(\mathbf{x}+\mathbf{y}) = a\mathbf{x} + a\mathbf{y}. \tag{2.8b}$$

The set of all vectors generated from a single vector in $\mathcal{V}$ by multiplication by different scalars, is called a *ray* in $\mathcal{V}$. Geometrically, a ray is represented by a line in the vector space.

Examples of Linear Vector Spaces

(1) The set of all n-tuples of numbers, $(x_1, x_2 \ldots x_n)$, when addition of vectors and multiplication of a vector by a scalar are defined by $(x_1, x_2 \ldots x_n) + (y_1, y_2 \ldots y_n) = (x_1 + y_1, x_2 + y_2 \ldots x_n + y_n)$, $a(x_1, x_2 \ldots x_n) = (ax_1, ax_2 \ldots ax_n)$.
This space is referred to as the n-dimensional *Euclidean* space.

(2) The real numbers, when they are considered both as vectors and as scalars. This is an example of a vector space consisting of a single ray, since all the vectors are generated from one vector (the number, 1) by multiplication by scalars.

For a general linear vector space, products of vectors (multiplication of a vector by a vector) need not be defined. However, we will restrict to spaces in which an *inner*, or *scalar*, *product* can be defined.

(v) A linear vector space is *unitary* if a scalar product is defined in it. That is, to every pair of vectors $\mathbf{x}$, $\mathbf{y}$, there corresponds a unique scalar (in general, complex), $(\mathbf{x}, \mathbf{y})$, called the *scalar product*, such that,

$$(\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x})^*, \tag{2.9a}$$

$$(\mathbf{x}, \mathbf{y}+\mathbf{z}) = (\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{z}), \tag{2.9b}$$

$$(\mathbf{x}, c\mathbf{y}) = c(\mathbf{x}, \mathbf{y}), \tag{2.9c}$$

$$(\mathbf{x}, \mathbf{x}) \geq 0, \text{ the equality sign holding only when } \mathbf{x} = \mathbf{0}. \tag{2.9d}$$

Here, the asterisk denotes complex conjugation. In $(\mathbf{x}, \mathbf{y})$, $\mathbf{x}$ is called the *prefactor* and $\mathbf{y}$ the *post-factor*. The scalar product is *linear* with respect to the post-factor:

$$(\mathbf{x}, a\mathbf{y} + b\mathbf{z}) = a(\mathbf{x}, \mathbf{y}) + b(\mathbf{x}, \mathbf{z}), \tag{2.10a}$$

and *antilinear* with respect to the prefactor.

$$(a\mathbf{x} + \mathbf{y}, \mathbf{z}) = a^*(\mathbf{x}, \mathbf{z}) + b^*(\mathbf{y}, \mathbf{z}) \tag{2.10b}$$

Because of this difference, $(\mathbf{x}, \mathbf{y})$ is sometimes called the scalar product of $\mathbf{y}$ by $\mathbf{x}$.

The 3-dimensional physical space (of position vectors) is an example of a unitary pace, while the 4-dimensional space-time of the theory of relatively ($x_1 = x, x_2 = y, x_3 = z, x_4 = ict$) is a linear vector space which is not unitary since $(\mathbf{x}, \mathbf{x}) \gtrless 0$, where $\mathbf{x}$ is a vector in a the space.

Norm of a vector: We define the *distance* between two vectors $\mathbf{x}$ and $\mathbf{y}$ by

$$||\mathbf{x}-\mathbf{y}|| \equiv +((\mathbf{x}-\mathbf{y}),(\mathbf{x}-\mathbf{y}))^{1/2}. \tag{2.11}$$

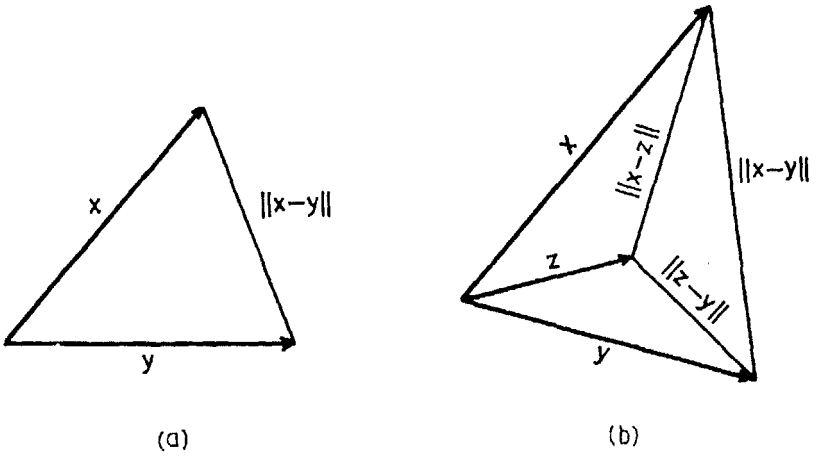


Fig. 2.1. Graphical representations of (a) Eq. (2.11), (b) semi-equality (2.12c).

Obviously,

$$||\mathbf{x}-\mathbf{y}|| = ||\mathbf{y}-\mathbf{x}||, \tag{2.12a}$$

$$||a\mathbf{x}-a\mathbf{y}|| = |a|\,||\mathbf{x}-\mathbf{y}||. \tag{2.12b}$$

$$||\mathbf{x}-\mathbf{y}|| \leq ||\mathbf{x}-\mathbf{z}|| + ||\mathbf{z}-\mathbf{y}||, \tag{2.12c}$$

$$||\mathbf{x}-\mathbf{x}|| = 0. \tag{2.12d}$$

The *norm* of $\mathbf{x}$ is the distance from the null vector (origin):

$$\text{Norm}\,(\mathbf{x}) \equiv ||\mathbf{x}|| = ((\mathbf{x}-\mathbf{0}),(\mathbf{x}-\mathbf{0}))^{1/2}$$

$$= (\mathbf{x}, \mathbf{x})^{1/2} \geq 0, \tag{2.13}$$

the last result following from the property of the scalar product. The norm of a vector in a unitary space, thus, corresponds to the length of a vector in the physical space. We note that the distance between $\mathbf{x}$ and $\mathbf{y}$ is, actually, the length of the vector $\mathbf{z} = \pm(\mathbf{x}-\mathbf{y})$.

Since the norm of a vector is zero only when the vector itself is zero, the norm of any non-zero vector is *positive definite*[2]. This property of the norm can also be

2. A quantity is positive definite if itself and its reciprocal are both positive. A real, positive number is necessarily positive definite. A purely imaginary number is an example of a quantity which could be positive without being positive definite.

expressed in the form of an inequality for the scalar product of any two vectors **x** and **y**:

$$|(\mathbf{x}, \mathbf{y})| \leq ||\mathbf{x}|| \cdot ||\mathbf{y}||, \tag{2.14}$$

where, the equality sign holds when $\mathbf{y} = a\mathbf{x}$. This is known as the *Schwarz inequality*. Thus, if θ_{xy} is the 'angle' between any two vectors **x** and **y**, we have,

$$|\cos \theta_{xy}| = \frac{(\mathbf{x}, \mathbf{y})}{||\mathbf{x}|| \cdot ||\mathbf{y}||} \leq 1. \tag{2.15}$$

Problem 2.1 : Prove the Schwarz inequality[3].

Orthonormality and Linear Independence

A vector, for which the norm is unity, is called a *unit vector*. From any given non-zero vector, a unit vector can be formed by dividing the given vector by its norm. Thus, $\mathbf{u} = \frac{\mathbf{x}}{||\mathbf{x}||}$, is a unit vector. **u** is then said to be *normalized*, and the process of forming **u** from **x** is called *normalization*.

Two vectors, **x** and **y**, are *orthogonal* if their scalar product is zero; that is, if

$$(\mathbf{x}, \mathbf{y}) = 0.$$

The unit vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_N$ form an *orthonormal set* if they are mutually orthogonal, i.e., if

$$(\mathbf{u}_i, \mathbf{u}_j) = \delta_{ij}. \tag{2.16}$$

The vectors belonging to the set $\mathbf{u}_1, \ldots, \mathbf{u}_N$ are *linearly independent* if none of them can be expressed as a linear combination of the others. Mathematically, this means that the equation,

$$\sum_{j=1}^{N} c_j \mathbf{u}_j = \mathbf{0}. \tag{2.17}$$

cannot be satisfied except by $c_j = 0$ for all j. For, suppose it is possible to satisfy Eq. (2.17) for non-zero values of c_j. Then, dividing the equation by c_i ($\neq 0$), we have,

$$\mathbf{u}_i = \sum_{j \neq i} b_j \mathbf{u}_j,$$

{where $b_j = -(c_j/c_i)$}, which contradicts our original statement that $\mathbf{u}_i$ cannot be expressed as a linear combination of the other $\mathbf{u}_j$'s. The only way to avoid the difficulty is to assume that $c_j = 0$, for $j \neq i$, so that, $c_i \mathbf{u}_i = 0$. Since $\mathbf{u}_i$ is a non-null vector, this requires that c_i also be zero.

The vectors belonging to an orthonormal set are necessarily linearly independent. The converse is, however, not true. But it is always possible to orthonormalize a set of linearly independent vectors. By this, we mean that, from a given

---

3. A proof is given in Merzbacher, E. *Quantum Mechanics* (John Wiley, New York 1970), II Edition, p. 298.

set of N linearly independent vectors, it is possible to form a set of N orthonormal vectors.

We have already shown how to normalise a given vector. Therefore, it would be sufficient for us to show how to form a set of orthogonal vectors from a given set of linearly independent unit vectors.

Let $\{\mathbf{v}_i\}_N$ denote the linearly independent set consisting of N unit vectors, such that,

$$(\mathbf{v}_i, \mathbf{v}_i) = 1; (\mathbf{v}_i, \mathbf{v}_j) \neq 0, \; i \neq j.$$

We will denote the orthogonal vectors by $\mathbf{u}_i$ $(i = 1, 2, \ldots N)$.

Let, $$\mathbf{u}_1 = \mathbf{v}_1 = a_{11}\mathbf{v}_1.$$

Since the given vectors are linearly independent,

$$\mathbf{v}_i \neq \sum_{j \neq i} c_j \mathbf{v}_j$$

Therefore,

$$\mathbf{u}_2 = \mathbf{v}_2 - (\mathbf{v}_1, \mathbf{v}_2)\mathbf{v}_1 = a_{22}\mathbf{v}_2 + a_{21}\mathbf{v}_1,$$

is orthogonal to $\mathbf{v}_1$ and, hence, to $\mathbf{u}_1$.

Similarly,

$$\mathbf{u}_n = \sum_{j=1}^{n} a_{nj}\mathbf{v}_j, \tag{2.18}$$

with

$$a_{nj} = -\sum_{k=ji=1}^{n-1\;\;k} a_{kj}a_{ki}^{*}(\mathbf{v}_i, \mathbf{v}_n), \; j < n; \tag{2.19a}$$

and

$$a_{nn} = 1, \tag{2.19b}$$

is orthogonal to $\mathbf{u}_i$ $(i = 1 \text{ to } n-1)$.

Eq. (2.18), for $n = 1$ to N, can be written as a matrix equation:

$$\mathbf{u} = A\mathbf{v}, \tag{2.20}$$

$$\mathbf{u} \equiv \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ . \\ . \\ \mathbf{u}_N \end{pmatrix}; \; \mathbf{v} \equiv \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ . \\ . \\ \mathbf{v}_N \end{pmatrix} \tag{2.21a}$$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ a_{21} & 1 & 0 & 0 & 0 & \ldots & 0 \\ a_{31} & a_{32} & 1 & 0 & 0 & \ldots & 0 \\ . & & & & & & \\ . & & & & & & \\ . & & & & & & \\ a_{N1} & a_{N2} & . & . & . & . & 1 \end{pmatrix} \tag{2.21b}$$

This process of forming orthogonal vectors from a given linearly independent set, is known as the *Schmidt* or the *Gram-Schmidt* orthogonalization procedure.

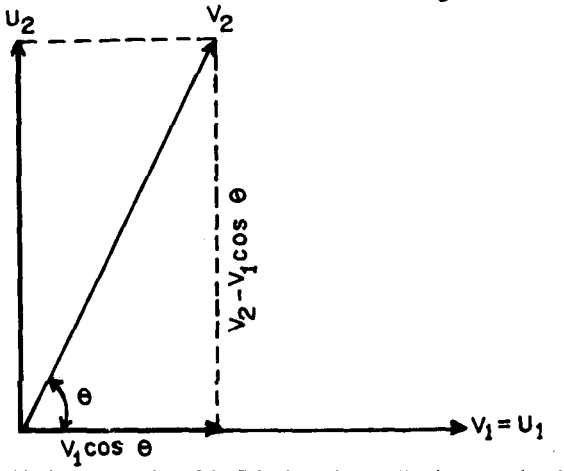


Fig. 2.2. Graphical representation of the Schmidt orthogonalization procedure in the case of two linearly independent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

As an illustrative example, let us consider a 2-dimensional space. Any two non-parallel vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, since $\mathbf{v}_1 \neq c\,\mathbf{v}_1$. We assume that $||\mathbf{v}_1|| = ||\mathbf{v}_2|| = 1$. Then,

$$\mathbf{u}_1 = \mathbf{v}_1$$

$$\mathbf{u}_2 = \mathbf{v}_2 - (\mathbf{v}_1, \mathbf{v}_2)\mathbf{v}_1$$

$$= \mathbf{v}_2 - \mathbf{v}_1 \cos\theta.$$

$$||\mathbf{u}_2|| = (1 - \cos^2\theta)^{1/2}.$$

Problem 2.2 : Show that only N orthogonal vectors can be formed from a given set of N linearly independent vectors.

Bases and Dimensions

A set $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_N$ of vectors is said to *span* a linear vector space if each vector in the space is a linear superposition, $\sum_{i=1}^{N} a_i \mathbf{v}_i$, of the elements of the set. A *basis* for a linear vector space is a set of linearly independent vectors that spans the space. Of course, there is an infinite number of bases in a given vector space. But the expansion of a vector in terms of a given basis is unique.

Let $\{\mathbf{v}_i\}_N$ be a basis, and let

$$\mathbf{X} = \sum_i x_i \mathbf{v}_i, \text{ and } \mathbf{X} = \sum_i x_i' \mathbf{v}_i,$$

be two different expansions of $\mathbf{X}$ in terms of $\{\mathbf{v}_i\}_N$. Then, from $\mathbf{X} - \mathbf{X} = \mathbf{0}$, we have,

$$(x_i - x_i')\mathbf{v}_i = \mathbf{0}.$$

Since the $\mathbf{v}_i$'s are linearly independent, this is possible only if

$$x_i - x_i' = 0, \text{ or, } x_i = x_i'.$$

A basis in the case of a linear vector space plays the role of a co-ordinate system in the case of the physical space, so that the expansion coefficient x_i could be regarded as the *component* of $\mathbf{X}$ along $\mathbf{v}_i$.

If the elements of a basis are orthonormal, we have an *orthonormal basis*. It is advantageous to use an orthonormal basis since, in this case, the expansion coefficient x_k in the expansion,

$$\mathbf{X} = \sum_i x_i \mathbf{u}_i, \tag{2.22}$$

can be found just by taking the scalar product of $\mathbf{X}$ by $\mathbf{u}_k$.

$$(\mathbf{u}_k, \mathbf{X}) = \sum_i x_i (\mathbf{u}_k, \mathbf{u}_i)$$

$$= \sum_i x_i \delta_{ik} = x_k. \tag{2.23}$$

Then, the product of two vectors $\mathbf{X}$ and $\mathbf{Y}$ is given by,

$$(\mathbf{X}, \mathbf{Y}) = \left(\sum_i x_i \mathbf{u}_i, \sum_k y_k \mathbf{u}_k\right)$$

$$= \sum_{i,k} x_i^* y_k (\mathbf{u}_i, \mathbf{u}_k) \tag{2.24}$$

$$= \sum_i x_i^* y_i.$$

Also,

$$(\mathbf{X}, \mathbf{X}) = \sum_i |x_i|^2 \geq 0, \tag{2.25}$$

as required by Eq. (2.9d).

In the following, we will denote an orthonormal basis by the symbol $[\ ]_N$.

The number N of elements in a basis $\{\mathbf{v}_i\}_N$ gives the *dimension* of the space. N may be finite or infinite. In a finite-dimensional space, every basis has the same number of elements. Also, any linearly independent set of N unit vectors would form a basis. These properties are not shared by infinite-dimensional spaces: Any linearly independent set having infinite number of elements is not a basis in such a space. Infinite-dimensional spaces have also other properties peculiar to themselves, which will be discussed later under Hilbert spaces.

Completeness (Closure Property)

A set $[\mathbf{u}_i]_N$ of orthonormal vectors in a linear vector space is *complete* if any other vector $\mathbf{X}$ in the space can be expanded in terms of the elements of the set (that is, if the set spans the space):

$$\mathbf{X} = \sum_{i=1}^{N} \mathbf{u}_i x_i, \tag{2.26}$$

where $x_i = (\mathbf{u}_i, \mathbf{X})$.

This means that the only vector that is orthogonal to all the $\mathbf{u}_i$'s is the null vector. A complete orthonormal set, thus, forms a basis (valid both for finite and infinite-dimensional spaces).

From Eq. (2.26), we have,

$$x_k = (\mathbf{u}_k, \mathbf{X}),$$

so that

$$\mathbf{X} = \sum_{i=1}^{N} \mathbf{u}_i(\mathbf{u}_i, \mathbf{X}). \tag{2.27}$$

Eq. (2.27) could be regarded as the criterion for the completeness of the set $[\mathbf{u}_i]_N$.

Since in a space of finite dimension N, the maximum number of vectors in a linearly independent set is N, the maximum number of vectors in an orthonormal set is also N, according to problem 2.2. Moreover, every orthonormal set containing N vectors is complete, and there should exist at least one orthonormal set that is complete.

As an illustration, let us consider the 3-dimensional physical space. A vector $\mathbf{V}$ in this space can be written as

$$\mathbf{V} = V_x\mathbf{i} + V_y\mathbf{j} + V_z\mathbf{k},$$

where, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the unit vectors along x, y and z directions, respectively. Then, corresponding to Eq. (2.27), we have,

$$\mathbf{V} = \mathbf{i}(\mathbf{i}\cdot\mathbf{V}) + \mathbf{j}(\mathbf{j}\cdot\mathbf{V}) + \mathbf{k}(\mathbf{k}\cdot\mathbf{V}),$$

which shows that $\mathbf{i}, \mathbf{j}, \mathbf{k}$ form a complete set. On the other hand, if our space is the xy-plane and $\mathbf{V}$ is a vector in this space, then,

$$\mathbf{V} = \mathbf{i}(\mathbf{i}\cdot\mathbf{V}) + \mathbf{j}(\mathbf{j}\cdot\mathbf{V}),$$

so that $\mathbf{i}, \mathbf{j}$ alone form a complete set of vectors in this space.

In order to extend the concept of completeness to a linear vector space, we have to introduce certain concepts and definitions concerning the convergence of a sequence.

x is the *limit point* of the sequence $\{x_n\}$ of (real or complex) numbers x_n $(n = 1, 2, \ldots\infty)$ if $|x - x_n| \to 0$ as $n \to \infty$. The sequence is then said to coverage to the limit x: $\{x_n\} \to x$. The limit point of a convergent sequence is unique: if $\{x_n\} \to x$ and $\{x_n\} \to x'$, then, $x - x' = 0$. But the limit point of a sequence need not be a member of the sequence.

Example: The sequence $\left\{\frac{1}{n}\right\} \equiv \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \ldots$ converges to the limit 0. But 0 is not a member of the sequence.

The sequence $\{x_n\}$ is a *Cauchy sequence* if $|x_n - x_m| \to 0$ as $n, m \to \infty$.

Every sequence which converges to a limit is a Cauchy sequence. For, if $\{x_n\} \to x$, then [Eq. (2.12c)],

$$|x_n - x_m| = |x_n - x + x - x_m| \leq |x_n - x| + |x - x_m| \to 0$$

as $n, m \to \infty$. Conversely, every Cauchy sequence should converge to a limit: From

$$|x_n - x| \leq |x_n - x_m| + |x_m - x|,$$

we see that $|x_n - x| \to |x_m - x|$ as $n, m \to \infty$, so that $|x_n - x|$ becomes independent of n as $n \to \infty$. This is possible only if x_n tends to some number x, which is not in the sequence, as $n \to \infty$; that is, $|x_n - x| \to 0$ as $n \to \infty$, which proves the statement.

The sequence $\left\{\frac{1}{n}\right\}$ quoted above is an example of a Cauchy sequence.

The limit point, or *limit vector*, of a sequence of vectors and a Cauchy sequence of vectors are similarly defined by replacing the number x_n by the vector $\mathbf{x}_n$ and the absolute value $|x_n - x_m|$ by the norm $||\mathbf{x}_n - \mathbf{x}_m||$ in the foregoing. An example of a Cauchy sequence of vectors is the sequence $\{\psi_n\}$ of partial sums,

$\psi_n = \sum_{k=1}^{n} a_k \phi_k$, of *square-integrable*, orthonormal functions[4] ϕ_k (where the a_k are numbers), such that $||\psi_n||^2 \equiv \sum_{k=1}^{n} |a_k|^2 < \infty$. The limit vector of this sequence is the function $\psi = \sum_{k=1}^{\infty} a_k \phi_k$. We are now in a position to state the condition for the completeness of a linear vector space:

A linear vector space is *complete* if every Cauchy sequence of vectors in the space converges to a limit vector *which is also in the space.*

Every finite-dimensional unitary space is necessarily complete. For, let $[\mathbf{u}_k]_N$ be an orthonormal basis for an N-dimensional space; and let $\{\mathbf{x}_m\}$ be a Cauchy sequence of vectors, where, $\mathbf{x}_m = \sum_{k=1}^{N} a_k^m \mathbf{u}_k$. Then,

$$||\mathbf{x}_m - \mathbf{x}_l||^2 = \sum_{k=1}^{N} |a_k^m - a_k^l|^2 \to 0,$$

as $m, l \to \infty$. That is $|a_k^m - a_k^l| \to 0$ as $m, l \to \infty$.

Thus, $\{a_k^m\}$ is a Cauchy sequence of numbers and, therefore, should converge to a limit, say a_k. Therefore, $x_m \to \mathbf{x} \equiv \sum_{k=1}^{N} a_k \mathbf{u}_k$, as $m \to \infty$. But $\mathbf{x}$, being a linear

---

4. The function $\phi_k(x)$ of a continuous variable x is said to be square-integrable if the Lebesque integral

$$\int_a^b |\phi_k(x)|^2 dx < \infty,$$

where $a \leq x \leq b$. The functions are orthonormal if

$$\int_a^b \phi_k^*(x)\phi_l(x)dx = \delta_{kl}.$$

The norm of ϕ_k is defined by

$$||\phi_k||^2 \equiv \int_a^b |\phi_k(x)|^2 dx.$$

superposition of the unit vectors $\mathbf{u}_k$, is a vector in the space. Thus, every Cauchy sequence of vectors converges to a vector which is in the space.

Hilbert Spaces

A Hilbert space is a unitary space that is complete. The Hilbert space is said to be *separable* if it contains a denumerable *dense* set of vectors[5]. In this case, every vector in the space can be approximated to a vector in this dense set.

An orthonormal basis $[\mathbf{u}_k]_N$, with k discrete, obviously, consists of a denumerable (finite or infinite) set of vectors, since we can form the basis starting from a single vector and then finding all the vectors that are orthonormal to it. This basis is also dense in the space since every vector $\mathbf{x}$ in the space can be written as $\mathbf{x} = \sum_{k=1}^{N} a_k \mathbf{u}_k$, which is the limit of the sequence $\{\mathbf{x}_n\}$ of partial sums $\mathbf{x}_n = \sum_{k=1}^{n} a_k \mathbf{u}_k$, where N is finite or infinite and $n \leq N$. Therefore, a Hilbert space is separable if it has an orthonormal basis[6].

We have seen that a finite-dimensional unitary space has an orthonormal basis and is complete. Therefore, every finite-dimensional unitary space is a separable Hilbert space. This space (finite-dimensional Hilbert space) is actually *isomorphic* to the n-dimensional Euclidean space; that is, there is a one-to-one correspondence between the vectors of an n-dimensional Hilbert space and the vectors of the n-dimensional Euclidean space.

An infinite-dimensional unitary space is not necessarily complete and, therefore, need not be a Hilbert space. However, only Hilbert spaces are of interest to quantum mechanics and, therefore, to us in this book.

An example of an infinite-dimensional Hilbert space is the space of infinite sequences of numbers $(x_1, x_2, \ldots x_k \ldots)$ such that $\sum_{k=1}^{\infty} |x_k|^2$ is finite, with addition, scalar multiplication and scalar product defined by :

---

5. A set of vectors is said to be *dense* in a linear vector space if every vector in the space is a limit point of the set (that is, limit point of a sequence of vectors in the set). Thus, every real number is the limit point of a sequence of rational numbers, so that the rational numbers constitute a dense set in the space of real numbers.
6. An example of a nondenumerable set of vectors is provided by the infinite set of momentum eigenfunctions $\mathbf{u}_\mathbf{k}(\mathbf{r})$ of a free particle (where $\hbar\mathbf{k}$ is the momentum of the particle):

$$\mathbf{u}_\mathbf{k}(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}} \exp(i\mathbf{k}\cdot\mathbf{r})$$

We have

$$(\mathbf{u}_\mathbf{k}, \mathbf{u}_{\mathbf{k}'}) \equiv \int \mathbf{u}_\mathbf{k}^*(\mathbf{r})\mathbf{u}_{\mathbf{k}'}(\mathbf{r})d^3\mathbf{r}$$

$$= \delta(\mathbf{k}-\mathbf{k}'), \qquad \text{[orthonormality]}$$

$$\int \mathbf{u}_\mathbf{k}^*(\mathbf{r})\mathbf{u}_\mathbf{k}(\mathbf{r}')d^3\mathbf{k} = \delta(\mathbf{r}-\mathbf{r}'), \qquad \text{[completeness]}$$

The nondenumerability arises from the fact that $\mathbf{k}$ is a continuous variable.

$$\mathbf{x}+\mathbf{y}=(x_1+y_1, x_2+y_2, \ldots, x_k+y_k, \ldots), \tag{2.28a}$$

$$a\mathbf{x}=(ax_1, ax_2, \ldots, ax_k, \ldots) \tag{2.28b}$$

$$(\mathbf{x}, \mathbf{y})=\sum_{k=1}^{\infty} x_k^* y_k, \tag{2.28c}$$

where $\mathbf{x} \equiv (x_1, x_2, \ldots, x_k \ldots)$ and

$\mathbf{y} \equiv (y_1, y_2, \ldots, y_k \ldots)$.

This space is called l^2. It is a separable Hilbert space since it has an orthonormal basis consisting of the vectors $\mathbf{u}_1=(1,0,0,\ldots), \mathbf{u}_2=(0,1,0,0\ldots)$ and so on, such that the only nonzero number in $\mathbf{u}_k$ is x_k which is equal to 1. An arbitrary vector $\mathbf{x}$ in l^2 can be written as $\mathbf{x}=\sum_{k=1}^{\infty} x_k \mathbf{u}_k$. The sequence of vectors $\{\mathbf{x}_n\}$, where $\mathbf{x}_n=\sum_{k=1}^{n} x_k^n \mathbf{u}_k$, is a Cauchy sequence whose limit vector is $\mathbf{x}$.

Just as every finite-dimensional Hilbert space is isomorphic to a Euclidean space of the same dimension, every infinite-dimensional separable Hilbert space is isomorphic to l^2.

Another example of an infinite-dimensional, separable Hilbert space is the space $L^2(a, b)$ of square integrable function $\phi(x)$ of a real variable x, with vector addition, scalar multiplication and scalar product defined as :

$$(\psi+\phi)(x)=\psi(x)+\phi(x),$$

$$(c\phi)(x)=c\phi(x)$$

$$(\psi, \phi)=\int_a^b \psi^*(x)\phi(x)dx.$$

The range a-b could be finite or infinite. The definition can also be extended to the case of functions of several independent variables.

Problem 2.3: Show that the functions 1, $\sqrt{2}\cos 2\pi kx$ and $\sqrt{2}\sin 2\pi kx$, for $k = 1, 2, 3\ldots\infty$, and $0 \le x \le 1$, constitute an orthonormal basis for the space $L^2(0, 1)$.

Suppose[7] $[\phi_k]_\infty$ is an orthonormal basis for an infinite-dimensional, separable Hilbert space. That is,

$$(\phi_k, \phi_l)=\delta_{kl}. \tag{2.29}$$

Then, any vector ψ in the space can be written as a linear combination of the ϕ_k's :

$$\psi=\sum_{k=1}^{\infty} a_k \phi_k \tag{2.30}$$

---

7. In the case of infinite-dimensional spaces, we will use the Greek alphabets ϕ, ψ and χ, in addition to bold face Latin alphabets, to represent vectors.

The coefficients a_k in (2.30) are unique, and are given by

$$a_l = \sum_{k=1}^{\infty} a_k(\phi_l, \phi_k) = (\phi_l, \psi), \tag{2.31}$$

according to (2.29). Thus, (2.30) becomes,

$$\psi = \sum_{k=1}^{\infty} \phi_k(\phi_k, \psi) \tag{2.32}$$

If the vectors ϕ_k are square-integrable functions, say of a variable x, then, Eqs. (2.29) and (2.32) can be written as

$$(\phi_k, \phi_l) = \int \phi_k^*(x)\phi_l(x)dx = \delta_{kl}. \tag{2.33}$$

$$\psi(x) = \sum_k \int \phi_k(x)\phi_k^*(x')\psi(x')dx'. \tag{2.34}$$

But

$$\psi(x) = \int \delta(x - x')\psi(x')dx', \tag{2.35}$$

where $\delta(x - x')$ is the *Dirac delta function* (see Appendix D). From (2.35) and (2.34) we see that

$$\sum_{k=1}^{\infty} \phi_k(x)\phi_k^*(x') = \delta(x - x'). \tag{2.36}$$

Eqs. (2.33) and (2.36) express, respectively, the orthonormality and the completeness of a set of vectors in the space L^2. These are of great importance for quantum mechanics because the wave function-space of the Schrödinger formulation of quantum mechanics is actually the space $L^2(-\infty, +\infty)$.

Linear Manifolds and Subspaces

A *linear manifold* in a linear vector space is a subset of vectors which itself forms a linear vector space. Thus, an n-dimensional Euclidean space is a linear manifold in the space l^2. The set of positive integers is an infinite-dimensional linear manifold in the space of real numbers. The set of vectors **i**, **j** constitutes a linear manifold in the 3-dimensional physical space.

A *subspace* $\mathcal{M}$ of a linear vector space is a *closed* linear manifold. Every finite-dimensional linear manifold is closed (as every finite-dimensional linear vector space is) and is, therefore, a subspace. However, infinite-dimensional spaces can have infinite-dimensional linear manifolds which are not necessarily closed.

A subspace of a separable Hilbert space is a separable Hilbert space. Two subspaces $\mathcal{M}_1$ and $\mathcal{M}_2$ of $\mathcal{H}$ are said to be orthogonal to each other if every vector in $\mathcal{M}_1$ is orthogonal to every vector in $\mathcal{M}_2$. The xy-plane and the z-axis are orthogonal subspaces of the physical space. Every separable Hilbert space can be split up into a finite number of subspaces $\mathcal{M}_1, \mathcal{M}_2 \ldots \mathcal{M}_n$ which are orthogonal to each other such that every vector ψ in $\mathcal{H}$ is the sum of n vectors, one each from each of the subspaces :

$$\psi = \sum_{k=1}^{n} \phi_k \tag{2.37}$$

where ϕ_k is a vector in $\mathcal{M}_k$. $\mathcal{H}$ is then said to be the *direct sum* of the $\mathcal{M}_k$'s : and is written as

$$\mathcal{H} = \mathcal{M}_1 \oplus \mathcal{M}_2 \oplus \ldots \oplus \mathcal{M}_n = \sum_{k=1}^{n} \oplus \mathcal{M}_k. \tag{2.38}$$

The vector ϕ_k is said to be the *projection* of ψ onto $\mathcal{M}_k$. If the set of vectors $\chi_1, \chi_2 \ldots$ forms an orthonormal basis for $\mathcal{M}_k$, then,

$$\phi_k = \sum_j (\chi_j, \psi)\chi_j. \tag{2.39}$$

As an illustrative example, let us consider the space $L^2(0, 1)$ of problem 2.3. The vectors $\chi_k = \sqrt{2} \sin 2\pi kx$ $(k = 1, 2, ...,)$ form an orthonormal basis for a subspace $\mathcal{M}_s$ [of $L^2(0, 1)$] in which an arbitrary vector is given by $\phi_s = \sum_{k=1}^{\infty} a_k \chi_k$. The set of vectors $\psi_0 = 1, \psi_k = \sqrt{2} \cos 2\pi kx$ $(k = 1, 2, \ldots)$ is an orthonormal basis for another subspace $\mathcal{M}_c$ which is orthogonal to $\mathcal{M}_s$. A vector in $\mathcal{M}_c$ is given by $\phi_c = \sum_{k=0}^{\infty} b_k \psi_k$. Since the sets $[\chi_k]_\infty$ and $[\psi_k]_\infty$ together form an orthonormal basis for $L^2(0, 1)$, any vector Ψ in $L^2(0, 1)$ can be written as $\Psi = \phi_s + \phi_c$, and $L^2(0, 1) = \mathcal{M}_s \oplus \mathcal{M}_c$.

2.2 OPERATORS

An *operator* on a linear vector space defines a (geometrical) relationship between two vectors. For example, if **Y** is obtained from **X** by rotating **X** about an axis, then the relationship between **X** and **Y** could be denoted by

$$\mathbf{Y} = \hat{A}\mathbf{X}, \tag{2.40}$$

where $\hat{A}$ is an operator[8] representing the rotation. We say that the result of operating with $\hat{A}$ on the vector **X**, is the vector **Y**.

Obviously; the operator $\hat{A}$ has meaning only with reference to a set of vectors. The space on which $\hat{A}$ *defined*, that is, the set of vectors **X** for which $\hat{A}\mathbf{X}$ has meaning, is called the *domain* of $\hat{A}$. The set of vectors **Y** expressible as $\mathbf{Y} = \hat{A}\mathbf{X}$, is called the *range* of $\hat{A}$. In a linear vector space if, $\mathbf{X}_1$ and $\mathbf{X}_2$ are in the domain of $\hat{A}$, then $(c_1\mathbf{X}_1 + c_2\mathbf{X}_2)$ is also in the domain of $\hat{A}$. An operator $\hat{A}$ is *linear* if

$$\hat{A}(c_1\mathbf{X}_1 + c_2\mathbf{X}_2) = c_1(\hat{A}\mathbf{X}_1) + c_2(\hat{A}\mathbf{X}_2), \tag{2.41a}$$

and *antilinear* if,

$$\hat{A}(c_1\mathbf{X}_1 + c_2\mathbf{X}_2) = c_1^*(\hat{A}\mathbf{X}_1) + c_2^*(\hat{A}\mathbf{X}_2). \tag{2.41b}$$

---

8. We use the symbol '^' in order to distinguish operators from scalars and vectors.

We will be concerned only with these two types of operators. The discussion in this section, however, will be confined to linear operators. A brief description of antilinear operators is given in Appendix B.

A linear operator preserves geometrical shapes ($PQ'R'$ in Fig. 2.3). An antilinear operator also preserves geometrical shapes, but it reflects the object on the real axis [$PQ''R''$ in Fig. 2.3].

The differential operator $\dfrac{d}{dx}$ is an example of a linear operator on the space $L^2(-\infty, \infty)$ of square integrable functions of a real variable x:

$$\frac{d}{dx}[c_1 f_1(x) + c_2 f_2(x)] = c_1 \frac{df_1}{dx} + c_2 \frac{df_2}{dx},$$

while $\hat{C}\dfrac{d}{dx}$, where $\hat{C}$ stands for complex conjugation, is an antilinear operator:

Example of a non-linear operator is the operator corresponding to squaring :

$$\mathrm{Sq}(c_1 x_1 + c_2 x_2) \neq \mathrm{Sq}(c_1 x_1) + \mathrm{Sq}(c_2 x_2).$$

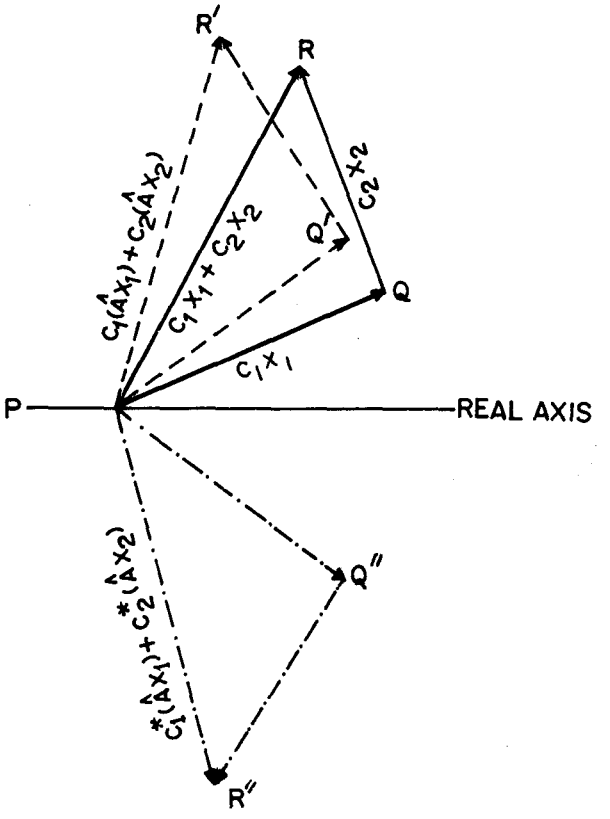


Fig. 2.3 Diagram showing the difference between a linear operator [Eq. (2.41a)] and an antilinear operator [Eq. (2.41b)].

The *null* (zero) and the *unit* (or, identity) operators are defined by the equations,

$$\hat{0}\mathbf{X} = \mathbf{0}, \tag{2.42}$$

and

$$\hat{1}\mathbf{X} = \mathbf{X}, \tag{2.43}$$

where $\mathbf{X}$ is any vector in the domain of these operators.

An algebra of linear operators can be constructed by defining the terms equality, sum, product, power, etc.

Equality: $\hat{A} = \hat{B}$, if the domains of $\hat{A}$ and $\hat{B}$ are identical, and if for every vector $\mathbf{X}$ in the domain,

$$\hat{A}\mathbf{X} = \hat{B}\mathbf{X}. \tag{2.44a}$$

$$\hat{A} > \hat{B}, \text{ if } (\mathbf{X}, \hat{A}\mathbf{X}) > (\mathbf{X}, \hat{B}\mathbf{X}), \tag{2.44b}$$

for all vectors $\mathbf{X}$ in the domain of $\hat{A}$ and $\hat{B}$.

Sum:

$$\hat{C} = \hat{A} + \hat{B},$$

if, for every vector $\mathbf{X}$ in the common domain of $\hat{A}$ and $\hat{B}$.

$$\hat{C}\mathbf{X} = \hat{A}\mathbf{X} + \hat{B}\mathbf{X}.$$

Product:

$$\hat{C} = \hat{A}\hat{B},$$

if for every vector $\mathbf{X}$ in the domain of $\hat{A}$ and $\hat{B}$,

$$\hat{C}\mathbf{X} = \hat{A}(\hat{B}\mathbf{X}).$$

In general $\hat{A}\hat{B} \neq \hat{B}\hat{A}$. That is, operators do not, in general, commute. Operator algebra is, therefore, said to be *non-commutative*. If $\hat{A}\hat{B} = \hat{B}\hat{A}$, $\hat{A}$ and $\hat{B}$ are called commuting operators[9].

To illustrate the noncommuting nature of operators, let us again consider the space of square-integrable functions of x. Let $\hat{A} \equiv x$, and $\hat{B} \equiv \dfrac{d}{dx}$. Then,

$$\hat{A}\hat{B}f(x) = x\frac{df}{dx}$$

$$\hat{B}\hat{A}f(x) = \frac{d}{dx}(xf) = \left(\frac{dx}{dx}\right)f + x\left(\frac{df}{dx}\right)$$

$$= f + x\frac{df}{dx}$$

9. The domain of $\hat{A}\hat{B}$ is a subspace of the domain of $\hat{B}$, whereas the domain of $\hat{B}\hat{A}$ is a subspace of the domain of $\hat{A}$. For, $\hat{A}\hat{B}\mathbf{X}$ is meaningful only when $\mathbf{X}$ is in the domain of $\hat{B}$ and $(\hat{B}\mathbf{X})$ is the domain of $\hat{A}$. For every vector $\mathbf{X}$ in the domain of $\hat{B}$, $\hat{B}\mathbf{X}$ may not be in the domain of $\hat{A}$ and, thus, may not be in the domain of $\hat{A}\hat{B}$. When $\hat{A}\hat{B} = \hat{B}\hat{A}$, however, the domains of $\hat{A}$, $\hat{B}$ and $\hat{A}\hat{B}$ coincide.

$$= f + \hat{A}\hat{B}f = (\hat{1} + \hat{A}\hat{B})f$$

Thus, $(\hat{A}\hat{B} - \hat{B}\hat{A})f(x) = -\hat{1}f(x)$,

or, $$x\frac{d}{dx} - \frac{d}{dx}x = -\hat{1},$$

or, $$\frac{d}{dx}x - x\frac{d}{dx} = \hat{1}.$$

Power: The square of $\hat{A}$ is defined as,

$$\hat{A}^2 = \hat{A}\hat{A}.$$

That is, $\hat{A}^2\mathbf{X} = \hat{A}(\hat{A}\mathbf{X})$, for every vector $\mathbf{X}$ in the domain of $\hat{A}$.
Similarly,

$$\hat{A}^n = \hat{A}\hat{A}^{n-1} = \underbrace{\hat{A}\hat{A}\ldots\hat{A}}_{n \text{ factors}}.$$

Function:

By combining the operations of addition and multiplication (product), a function of an operator can be formed.

Ex: $$\hat{F} \equiv a\frac{d^2}{dx^2} + b\frac{d}{dx} + c\hat{1},$$

is a function of the differential operator $\frac{d}{dx}$. Function of a linear operator is a linear operator. That is,

$$\hat{F}[c_1 f_1(x) + c_2 f_2(x)] = c_1(\hat{F}f_1) + c_2(\hat{F}f_2).$$

Inverse:

If two operators $\hat{A}$ and $\hat{B}$ are related by[10]

$$\hat{A}\hat{B} = \hat{B}\hat{A} = \hat{1}, \tag{2.45}$$

then, they are said to be reciprocal to each other and $\hat{B}$, denoted by $\hat{A}^{-1}$, is called the *inverse* of $\hat{A}$. An operator for which an inverse exists is said to be *non-singular*, whereas one for which no inverse exists is *singular*. A necessary and sufficient condition for an operator $\hat{A}$ to be non-singular is that corresponding to each vector $\mathbf{Y}$, there should be a unique vector $\mathbf{X}$ such that $\mathbf{Y} = \hat{A}\mathbf{X}$.

Problem 2.4: Prove the preceding statement. Hence show that, in the space of the square-integrable functions of the variable x, the operator $\hat{x}$ has an inverse, $\hat{0}$ has no inverse and $\hat{1}$ is inverse of itself.

Inverse of a linear operator, is a linear operator:

Let $$\mathbf{Y}_1 = \hat{A}\mathbf{X}_1, \mathbf{Y}_2 = \hat{A}\mathbf{X}_2.$$

---

10. Note that both the conditions in Eq. (2.45) are necessary for the existence of an inverse. However, in the case of finite-dimensional spaces, $\hat{B}\hat{A} = \hat{1}$ implies $\hat{A}\hat{B} = \hat{1}$.

Then, $$\mathbf{X}_1 = \hat{A}^{-1}\mathbf{Y}_1, \mathbf{X}_2 = \hat{A}^{-1}\mathbf{Y}_2,$$

so that $$c_1\mathbf{X}_1 = c_1(\hat{A}^{-1}\mathbf{Y}_1), c_2\mathbf{X}_2 = c_2(\hat{A}^{-1}\mathbf{Y}_2).$$

Thus, $$\hat{A}^{-1}[c_1\mathbf{Y}_1 + c_2\mathbf{Y}_2] = \hat{A}^{-1}[c_1(\hat{A}\mathbf{X}_1) + c_2(\hat{A}\mathbf{X}_2)]$$

$$= \hat{A}^{-1}\hat{A}[c_1\mathbf{X}_1 + c_2\mathbf{X}_2],$$

since $\hat{A}$ is linear.

$$= c_1\mathbf{X}_1 + c_2\mathbf{X}_2$$

$$= c_1(\hat{A}^{-1}\mathbf{Y}_1) + c_2(\hat{A}^{-1}\mathbf{Y}_2). \tag{2.46}$$

The inverse of a product of operators is the product of the inverse in the reverse order:

$$(\hat{A}\hat{B}\hat{C})^{-1} = \hat{C}^{-1}\hat{B}^{-1}\hat{A}^{-1}.$$

This could be shown as follows: We have

$$\hat{A}\hat{B}\hat{C}(\hat{A}\hat{B}\hat{C})^{-1} = \hat{1}.$$

Multiplying successively from the left by $\hat{A}^{-1}$, $\hat{B}^{-1}$ and $\hat{C}^{-1}$, we get

$$(\hat{A}\hat{B}\hat{C})^{-1} = \hat{C}^{-1}\hat{B}^{-1}\hat{A}^{-1}. \tag{2.47}$$

Eigenvalues and Eigenvectors of an Operator

The result of operating on a vector with an operator $\hat{A}$ is, in general, a different vector:

$$\hat{A}\mathbf{X} = \mathbf{Y}. \tag{2.48}$$

But there may be some vector $\mathbf{X}$ with the property,

$$\hat{A}\mathbf{X} = \alpha\mathbf{X}, \tag{2.49}$$

where, α is a scalar. That is, the operation with $\hat{A}$ on $\mathbf{X}$ yields the same vector $\mathbf{X}$ multiplied by a scalar. $\mathbf{X}$ is, then, called an *eigenvector* (or eigenfunction) of $\hat{A}$ belonging to the *eigenvalue* α. Eq. (2.49) is the *eigenvalue equation* for $\hat{A}$. A linear operator has, in general, several eigenvalues and eigenvectors, which are then distinguished by a subscript:

$$\hat{A}\mathbf{X}_k = \alpha_k\mathbf{X}_k.$$

The set $\{\alpha_k\}$ of all the eigenvalues taken together constitute the *spectrum* of the operator. The eigenvalues may be discrete, continuous or partly discrete and partly continuous. An eigenvector belongs to only one eigenvalue. But several linearly independent eigenvectors may belong to the same eigenvalue. In this case, the eigenvalue is said to be *degenerate*, and the number of linearly independent eigenvectors is the *degree of degeneracy*. The eigenvectors belonging to a degenerate eigenvalue of a linear operator, span a subspace whose dimensionality is equal to the degree of degeneracy.

Problem 2.5: Prove the above statement.

For an illustration, let us consider the operation $-\frac{d^2}{dx^2}$. The eigenvalue equation of this operator is (we write E instead of α and $\phi(x)$ in place of $\mathbf{X}$),

$$\left(-\frac{d^2}{dx^2}-E\right)\phi(x)=0.$$

The two linearly independent eigenvectors are

$$\phi_p(x)=e^{ipx},$$

and

$$\phi_{-p}(x)=e^{-ipx},$$

where $p^2=E$. Both the eigenvectors belong to the same eigenvalue E. Hence E is 2-fold degenerate. If p is treated as a continuous variable, then, the eigenvalue spectrum is continuous.

The eigenvalue of the square of an operator is the square of the eigenvalue of the operator. For, if $\hat{A}\mathbf{X}=\alpha\mathbf{X}$, we have,

$$\hat{A}^2\mathbf{X}=\hat{A}(\hat{A}\mathbf{X})=\alpha(\hat{A}\mathbf{X})=\alpha^2\mathbf{X}. \tag{2.50}$$

Operators of special importance in quantum mechanics are considered below:

(i) Positive Definite Operator

$\hat{A}$ is positive definite, if $\hat{A}>\hat{0}$ and $\hat{A}^{-1}>\hat{0}$. (ef. footnote 2). *The eigenvalues of a positive definite operator, are all positive.*

If $\hat{A}>\hat{B}>\hat{0}$, then, $\hat{B}^{-1}>\hat{A}^{-1}>0.$ (2.51)

For, $\hat{A}^{-1}\hat{A}\hat{A}^{-1}>\hat{A}^{-1}\hat{B}\hat{A}^{-1}.$

That is, $\hat{A}^{-1}(\hat{1}-\hat{B}\hat{A}^{-1})>\hat{0}.$

But, $\hat{B}\hat{A}^{-1}<\hat{A}\hat{A}^{-1}=\hat{1}.$

Therefore, $(\hat{1}-\hat{B}\hat{A}^{-1})>\hat{0}.$

Therefore, $\hat{A}^{-1}>\hat{0}.$

Now, $\hat{B}^{-1}\hat{A}>\hat{B}^{-1}\hat{B}=\hat{A}^{-1}\hat{A}=\hat{1}.$

Thus, $\hat{B}^{-1}>\hat{A}^{-1}>\hat{0}.$

Problem 2.6: If $\hat{A}$ and $\hat{B}$ are positive definite operators, show that $(\hat{A}\hat{B})$ is a positive definite operator.

(ii) Continuous and Bounded Operators

An operator $\hat{A}$ is *continuous* if $\{\hat{A}\mathbf{X}_n\} \to \hat{A}\mathbf{X}$, for every Cauchy sequence of vectors $\{\mathbf{X}_n\}$ that converges to the limit $\mathbf{X}$. $\hat{A}$ is *bounded* if there is a positive number 'a' such that

$$||\hat{A}\mathbf{X}|| \leq a||\mathbf{X}||,$$

for every vector $\mathbf{X}$ in the domain of $\hat{A}$. The smallest 'a' with this property is called the norm of $\hat{A}$, and is denoted by $||\hat{A}||$. Thus, $||\hat{A}||$ is defined by,

$$||\hat{A}\mathbf{X}||_{\max} = ||\hat{A}|| \cdot ||\mathbf{X}||. \tag{2.52}$$

$\hat{A}$ is continuous if, and only if, it is bounded:
For, $||\hat{A}\mathbf{X}_n - \hat{A}\mathbf{X}|| \leq ||\hat{A}|| \cdot ||\mathbf{X}_n - \mathbf{X}|| \to 0$, if $\hat{A}$ is bounded, and if $\{\mathbf{X}_n\} \to \mathbf{X}$. Thus, $\hat{A}\mathbf{X}_n \to \hat{A}\mathbf{X}$, for $n \to \infty$. If $\hat{A}$ is not bounded, there is a vector $\mathbf{X}_n$ corresponding to every positive integer n, such that $||\hat{A}\mathbf{X}_n|| > n||\mathbf{X}_n||$. Defining the vector $\mathbf{Y}_n$ by $\mathbf{Y}_n = \frac{1}{n||\mathbf{X}_n||} \cdot \mathbf{X}_n$, we have, $||\mathbf{Y}_n|| = \frac{1}{n}$, $\mathbf{Y}_n \to \mathbf{0}$ as $n \to \infty$, showing that $\{\mathbf{Y}_n\}$ is a Cauchy sequence that converges to the null vector. Therefore, $\hat{A}$ will be continuous, if $\hat{A}\mathbf{Y}_n \to 0$ as $n \to \infty$; that is, if $||\hat{A}\mathbf{Y}_n|| \to 0$. But $||\hat{A}\mathbf{Y}_n|| = \frac{1}{n||\mathbf{X}_n||} \cdot ||\hat{A}\mathbf{X}_n|| > 1$.

Thus $\hat{A}$ is not continuous.

The following properties of the norm of a bounded linear operator could easily be proved:

$$||\hat{A} + \hat{B}|| \leq ||\hat{A}|| + ||\hat{B}||, \tag{2.53a}$$

$$||c\hat{A}|| = |c| \cdot ||\hat{A}||, \tag{2.53b}$$

$$||\hat{A}\hat{B}|| \leq ||\hat{A}|| \cdot ||\hat{B}||, \tag{2.53c}$$

$$||\hat{A}|| = 0, \text{ if, and only if } \hat{A} = \hat{0}. \tag{2.53d}$$

Properties (2.53a, b, d) respectively, correspond to, and derive from, the properties (2.12c, b, d) of a vector in a linear vector space, whereas property (2.53c) derives from the definition of the norm and is the equivalent of the Schwarz inequality (2.14) satisfied by vectors. These properties show that the norm of a bounded operator has the characteristics of length. They also imply that the sums, products and scalar multiples of bounded operators are bounded.

Every operator defined on a finite-dimensional space, is bounded. This property is not shared by operators defined on infinite-dimensional spaces. Since, in quantum mechanics, we have to deal with infinite-dimensional spaces, we have to consider unbounded operators. An example of an unbounded operator is the operator $\hat{x}$ on the space $L^2(-\infty, \infty)$ of square integrable functions $\phi(x)$ of the real variable x. For,

$$||\hat{x}\phi||^2 = \int_{-\infty}^{+\infty} |\hat{x}\phi(x)|^2 \, dx,$$

which need not be a finite number times $||\phi|| \equiv \int_{-\infty}^{+\infty} |\phi(x)|^2 \, dx$.

(iii) Hermitian Operators

Two bounded linear operators $\hat{A}$ and $\hat{B}$ are said to be *adjoint* of each other if, for arbitrary vectors **X** and **Y** in the domain of $\hat{A}$ and $\hat{B}$ (assumed to be the same),

$$(\hat{A}\mathbf{X}, \mathbf{Y}) = (\mathbf{X}, \hat{B}\mathbf{Y}). \tag{2.54}$$

$\hat{B}$ is, then, denoted by $\hat{A}^\dagger$. $\hat{A}$ is *self-adjoint* or *Hermitian* if

$$\hat{A} = \hat{A}^\dagger, \tag{2.55a}$$

and anti-Hermitian if

$$\hat{A} = -\hat{A}^\dagger. \tag{2.55b}$$

Thus, Hermitian operators are defined by,

$$(\hat{A}\mathbf{X}, \mathbf{Y}) = (\mathbf{X}, \hat{A}\mathbf{Y}), \tag{2.56}$$

Unlike bounded operators, an unbounded operator (like $\hat{x}$, considered in a previous example) can be Hermitian only with respect to a restricted number of vectors (see, Ref. 1, Section 9).

Hermitian operators have the following important properties :

(*H*1). *The eigenvalues are real*:

Let $\hat{H}$ be the Hermitian operator and let **X** be an eigenvector belonging to the eigenvalue λ :

$$\hat{H}\mathbf{X} = \lambda\mathbf{X}.$$

By definition, we have,

$$(\hat{H}\mathbf{X}, \mathbf{X}) = (\mathbf{X}, \hat{H}\mathbf{X}).$$

That is [see Eqs. (2.9a, c)].

$$(\lambda^* - \lambda)(\mathbf{X}, \mathbf{X}) = 0.$$

Since $(\mathbf{X}, \mathbf{X}) \neq 0, \lambda^* = \lambda.$

(*H*2). *Eigenvectors belonging to different eigenvalues are orthogonal :*

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be eigenvectors of $\hat{H}$ belonging, respectively, to the eigenvalues λ_1 and λ_2:

$$\hat{H}\mathbf{X}_1 = \lambda\mathbf{X}_1; \hat{H}\mathbf{X}_2 = \lambda_2\mathbf{X}_2$$

Then, $$(\mathbf{X}_2, \hat{H}\mathbf{X}_1) - (\hat{H}\mathbf{X}_2, \mathbf{X}_1) = 0.$$

That is, $$(\lambda_1 - \lambda_2)(\mathbf{X}_2, \mathbf{X}_1) = 0, \text{ since } \lambda_2^* = \lambda_2$$

By hypothesis, $(\lambda_1 - \lambda_2) \neq 0$, so that,

$$(\mathbf{X}_2, \mathbf{X}_1) = 0.$$

(*H*3). *The set of all eigenvectors of a bounded Hermitian operator forms a complete set.* Since the eigenvectors are orthogonal, and since we can normalize them, this means that the eigenvectors form a basis for the space.

Problem 2.7: Show that the adjoint of a product of operators, is the product of the adjoints in the reverse order. Hence show that the product of two Hermitian operators, is Hermitian only if the operators commute (i.e., $\hat{A}\hat{B} = \hat{B}\hat{A}$).

Problem 2.8: If $\hat{H}$ is Hermitian, show that $(\mathbf{X}, \hat{H}\mathbf{X})$ is real, for any vector $\mathbf{X}$ in the domain of $\hat{H}$.

Problem 2.9: Show that, for the space $L^2(-\infty, \infty)$ of square-integrable functions of x, the operator $i\dfrac{d}{dx}$ is Hermitian whereas $\dfrac{d}{dx}$ is not.

Problem 2.10: Show that the operator, $\hat{B} \equiv i\hat{A}$ is anti-Hermitian when $\hat{A}$ is Hermitian.

(iv) Unitary Operators

A linear operator $\hat{U}$ is *unitary* if it preserves the Hermitian character of an operator under a *similarity transformation*. Now, a similarity transformation of an operator $\hat{A}$ by a *non-singular* operator $\hat{S}$, is defined as,

$$\hat{A} \rightarrow A' = \hat{S}\hat{A}\hat{S}^{-1}. \tag{2.57}$$

Thus, the condition for $\hat{U}$ to be unitary is that,

$$(\hat{U}\hat{A}\hat{U}^{-1})^{\dagger} = \hat{U}\hat{A}\hat{U}^{-1}. \tag{2.58}$$

where, $$\hat{A}^{\dagger} = \hat{A}.$$

But (see Prob. 2.7),

$$(\hat{U}\hat{A}\hat{U}^{-1})^{\dagger} = (\hat{U}^{-}1)^{\dagger}\hat{A}\hat{U}^{\dagger}, \text{ so that,}$$

$$(\hat{U}^{-1})^{\dagger}\hat{A}\hat{U}^{\dagger} = \hat{U}\hat{A}\hat{U}^{-1}.$$

Multiplying from the left by $\hat{U}^{\dagger}$ and from the right by $\hat{U}$, we get,

$$\hat{U}^{\dagger}(\hat{U}^{-1})^{\dagger}\hat{A}\hat{U}^{\dagger}\hat{U} = \hat{U}^{\dagger}\hat{U}\hat{A}.$$

That is, $$\hat{A}(\hat{U}^{\dagger}\hat{U}) = (\hat{U}^{\dagger}\hat{U})\hat{A}, \tag{2.59}$$

since $$\hat{U}^{\dagger}(\hat{U}^{-1})^{\dagger} = (\hat{U}^{-1}\hat{U})^{\dagger} = \hat{1}.$$

Now, only the identity operator has the property that

$$\hat{A}\hat{1} = \hat{1}\hat{A}.$$

for arbitrary operator $\hat{A}$. Hence,

$$\hat{U}^{\dagger}\hat{U} = \hat{1}. \tag{2.60}$$

In the case of infinite-dimensional spaces, Eq. (2.60), by itself, does not imply the condition that $\hat{U}$ should have an inverse. This condition can be incorporated by multiplying both sides of the equation by $\hat{U}^{-1}$. Then, the condition for the unitarity of $\hat{U}$ becomes,

$$\hat{U}^{\dagger} = \hat{U}^{-1} \tag{2.61a}$$

or,

$$\hat{U}\hat{U}^{\dagger} = \hat{U}^{\dagger}\hat{U} = \hat{1}. \tag{2.61b}$$

Under the operation of $\hat{U}$, a vector $\mathbf{X}$ is transformed into the vector $\mathbf{X}' = \hat{U}\mathbf{X}$. Thus, if two vectors $\mathbf{X}$ and $\mathbf{Y}$ are transformed by the same unitary operator $\hat{U}$, then,

$$(\mathbf{X}',\ \mathbf{Y}') = (\hat{U}\mathbf{X}, \hat{U}\mathbf{Y}) = (\mathbf{X}, \hat{U}^{\dagger}U\mathbf{Y})$$

$$= (\mathbf{X},\ \mathbf{Y}). \tag{2.62}$$

Thus, the transformation by a unitary operator (that is, a *unitary* transformation) preserves the scalar product of vectors. In particular, it leaves the norm of a vector unchanged. Now, a transformation that leaves both the lengths of vectors and the angles between vectors unchanged, is a rotation, Thus, a unitary transformation in a linear vector space, is analogous to a rotation in the physical space.

Since $||\hat{U}\mathbf{X}|| = ||\mathbf{X}||$ for every vector $\mathbf{X}$, $\hat{U}$ is bounded ($||\hat{U}|| = 1$).

Corresponding to every unitary operator $\hat{U}$, we can define a Hermitian operator $\hat{H}$ and *vice verse*, by

$$\hat{U} = \exp(i \epsilon \hat{H}) \tag{2.63}$$

where ϵ is a parameter. Obviously (see Eq. (2.55a) and Problem 2.10).

$$\hat{U}^{\dagger} = \exp[(i \epsilon \hat{H})]^{\dagger} = \exp(-i \epsilon \hat{H}) = \hat{U}^{-1}.$$

This means that every Hermitian operator is the *generator* of a unitary transformation.

The following properties of a unitary operator could be easily proved:

($U1$). The eigenvalues are unimodular. That is, if $\hat{U}\mathbf{X} = \alpha\mathbf{X}$, then $|\alpha| = 1$.

($U2$). Eigenvectors belonging to different eigenvalues are orthogonal.

($U3$). Product of two or more unitary operators are unitary.

Problem 2.11: Prove the above properties of the unitary operator.

(v) Projection Operators

Consider an operator $\hat{I}$ defined by,

$$\hat{I}^2 = \hat{1}. \tag{2.64}$$

The eigenvalues of $\hat{I}$ are, then, +1 and −1 [see, Eq. (2.50)]. Let the corresponding eigenvectors be $\mathbf{X}_+$ and $\mathbf{X}_-$, respectively:

$$\hat{I}\mathbf{X}_{\pm} = \pm\mathbf{X}_{\pm}. \tag{2.65}$$

$\hat{I}^2$ is, obviously, Hermitian, so that $\hat{I}$ is a Hermitian operator. Hence $\mathbf{X}_+$ and $\mathbf{X}_-$ are orthogonal to each other (Property *H*2 of Hermitian operators):

$$(\mathbf{X}_+, \mathbf{X}_-) = 0. \tag{2.66}$$

The operator $\hat{I}$ is called an *involution*.

Now, let us define the operators π_+ and π_- by

$$\hat{\pi}_{\pm} = \frac{1}{2}(\hat{1} \pm \hat{I}). \tag{2.67}$$

Then,

$$\hat{I}(\hat{\pi}_{\pm}\mathbf{X}) = \frac{1}{2}(\hat{I} \pm \hat{1})\mathbf{X} = \pm(\hat{\pi}_{\pm}\mathbf{X}),$$

where $\mathbf{X}$ is an arbitrary vector. This shows that, [see Eq. (2.65)],

$$\pi_{\pm}\mathbf{X} = \mathbf{X}_{\pm}. \tag{2.68}$$

Also, we have,

$$\hat{\pi}_{\pm}^2 = \hat{\pi}_{\pm}, \tag{2.69a}$$

or,

$$\hat{\pi}_{\pm}(\hat{\pi}_{\pm} - \hat{1}) = \hat{0},$$

so that,

$$\hat{0} \le \hat{\pi}_{\pm} \le \hat{1}. \tag{2.69b}$$

$$\hat{\pi}_{\pm}^{\dagger} = \hat{\pi}_{\pm}, \tag{2.69c}$$

$$\hat{\pi}_{+}\hat{\pi}_{-} = \hat{0} = \hat{\pi}_{-}\hat{\pi}_{+}, \tag{2.69d}$$

$$\hat{\pi}_{+} + \hat{\pi}_{-} = \hat{1}. \tag{2.69e}$$

From Eqs. (2.69e) and (2.68), we get,

$$\mathbf{X} = \mathbf{X}_{+} + \mathbf{X}_{-}, \tag{2.70}$$

whereas, Eq. (2.69d) shows that,

$$\hat{\pi}_{\pm}\mathbf{X}_{\mp} = 0. \tag{2.71}$$

Thus, every vector $\mathbf{X}$ is a linear sum of $\mathbf{X}_+$ and $\mathbf{X}_-$. If the set of vectors, $\{\mathbf{X}\}$, spans a linear vector space, then, the sets $\{\mathbf{X}_+\}$ and $\{\mathbf{X}_-\}$ span two distinct subspaces orthogonal to each other. We say that the original space is *reduced* by the Hermitian involution $\hat{I}$ into two subspaces orthogonal to each other, such that $\mathbf{X}$ is a unique linear combination of two vectors, one from each subspace. The vectors $\mathbf{X}_+$ and $\mathbf{X}_-$ are the *projections* of $\mathbf{X}$ onto the subspaces, and $\hat{\pi}_+$ and $\hat{\pi}_-$ are the respective *projection operators*.

Eqs. (2.69a) and (2.69c) together could be taken as the definition of a projection operator[11], whereas Eq. (2.69d) expresses the orthogonality of $\hat{\pi}_+$ and $\hat{\pi}_-$ (that is, the fact that they project onto subspaces orthogonal to each other). The unit operator $\hat{1}$ and the null operator $\hat{0}$ could be regarded as projection operators that project onto the whole space and onto the subspace containing only the null vector, respectively. A projection operator is, obviously, bounded since $||\hat{\pi}\psi|| \leq ||\psi||$ for every projection operator $\hat{\pi}$, so that $||\hat{\pi}|| = 1$, except for $\hat{0}$ for which $||\hat{0}|| = 0$.

Illustrative Example :

Consider a vector **V** in *xy*-plane. **V** can be written as (see Fig. 2.4),

$$\mathbf{V} = \mathbf{V}_x + \mathbf{V}_y,$$

where,

$$\mathbf{V}_x = V_x\mathbf{i} = (\mathbf{i}, \mathbf{V}),$$

$$\mathbf{V}_y = V_y\mathbf{j} = \mathbf{j}(\mathbf{j}, \mathbf{V}); \mathbf{i} \text{ and } \mathbf{j} \text{ being unit vectors.}$$

Writing $\mathbf{V} \equiv \mathbf{V})$, etc., we have[12],

$$\mathbf{V}_x) = \hat{\pi}_x\mathbf{V}) = \mathbf{i})(\mathbf{i}, \mathbf{V}),$$

$$\mathbf{V}_y) = \hat{\pi}_y\mathbf{V}) = \mathbf{j})(\mathbf{j}, \mathbf{V}).$$

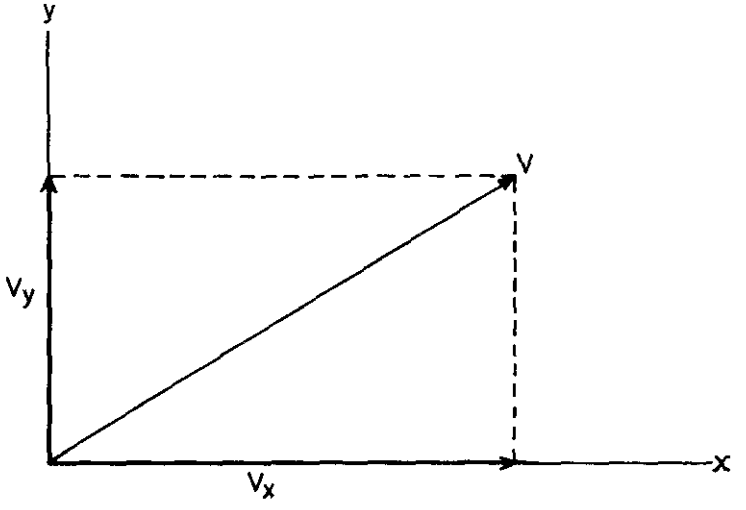


Fig. 2.4. The resolution of a vector into its projections onto orthogonal subspaces *x* and *y*.

Thus, the projection operators $\hat{\pi}_x$ and $\hat{\pi}_y$ are given by,

$$\hat{\pi}_x \equiv \mathbf{i})(\mathbf{i},$$

---

11. An operator obeying Eq. (2.69a) : $\hat{A}^2 = \hat{A}$, is called *idempotent*.
12. A more elegant notation for vectors, due to Dirac, will be discussed in the next section.

$$\hat{\pi}_y \equiv \mathbf{j})(\mathbf{j}.$$

If V and W are two vectors in the xy-plane, we have,

$$(\mathbf{V}, \hat{\pi}_x \mathbf{W}) = (\mathbf{V},\ \mathbf{i})(\mathbf{i},\ \mathbf{W}) = (\hat{\pi}_x \mathbf{V},\ \mathbf{W}),$$

since

$$(\hat{\pi}_x \mathbf{V}, \mathbf{W}) = (\mathbf{W}, \hat{\pi}_x \mathbf{V})^* = (\mathbf{W},\ \mathbf{i})^*(\mathbf{i},\ \mathbf{V})^*$$

$$= (\mathbf{i},\ \mathbf{W})(\mathbf{V},\ \mathbf{i}).$$

Thus, $\hat{\pi}_x$ is Hermitian.

$$\hat{\pi}_x^2 = \mathbf{i})(\mathbf{i},\ \mathbf{i})(\mathbf{i} =\ \mathbf{i})(\mathbf{i},\ \text{since } (\mathbf{i}, \mathbf{i}) = 1.$$

$$\hat{\pi}_x \hat{\pi}_y = \mathbf{i})(\mathbf{i},\ \mathbf{j})(\mathbf{i} = \mathbf{0},\ \text{since } (\mathbf{i}, \mathbf{j}) = 0.$$

Also,

$$[\hat{\pi}_x + \hat{\pi}_y]\mathbf{V}) = \mathbf{V}),$$

so that,

$$\hat{\pi}_x + \hat{\pi}_y = \hat{1}.$$

$\hat{\pi}_x$ and $\hat{\pi}_y$, thus, have properties identical to those of $\hat{\pi}_+$ and $\hat{\pi}_-$. The involution operator, in this case, is seen to be,

$$\hat{I} = \mathbf{i})(\mathbf{i} - \mathbf{j})(\mathbf{j}.$$

The foregoing considerations could be extended to the case where there are more than two projection operators. Let $[\mathbf{u}_i]_N$ be an orthonormal basis :

$$(\mathbf{u}_i, \mathbf{u}_j) = \delta_{ij};\ \Sigma_j \mathbf{u}_j)(\mathbf{u}_j = \hat{1}.$$

Then, we can define $\hat{\pi}_i$ by,

$$\hat{\pi}_i \mathbf{u}_j = \delta_{ij} \mathbf{u}_i,\ i = 1, 2, \ldots N. \tag{2.72}$$

$\hat{\pi}_i$ is the projection operator that projects onto the one-dimensional subspace defined by $\mathbf{u}_i$. An arbitrary vector X in the space can be written as, (see Eq. (2.26))

$$\mathbf{X}) = \sum_{j=1}^{N} \mathbf{u}_j)(\mathbf{u}_j, \mathbf{X}), \tag{2.73}$$

so that,

$$\hat{\pi}_i \mathbf{X}) = \Sigma_j \hat{\pi}_i \mathbf{u}_j)(\mathbf{u}_j, \mathbf{X}) = \mathbf{u}_i)(\mathbf{u}_i, \mathbf{X}).$$

Hence,

$$\hat{\pi}_i \equiv \mathbf{u}_i)(\mathbf{u}_i, \tag{2.74}$$

from which it is easily shown that

$$\hat{\pi}_i \hat{\pi}_j = \delta_{ij} \hat{\pi}_j. \tag{2.75}$$

Substituting Eq. (2.74) in Eq. (2.73), we get,

$$\mathbf{X}) = \sum_{j=1}^{N} \hat{\pi}_j \mathbf{X}),$$

so that,

$$\sum_{j=1}^{N} \hat{\pi}_j = \hat{1}. \tag{2.76}$$

Eq. (2.76) is the analogue of Eq. (2.69e), and is called the *resolution of the identity* in terms of the projection operators. It is actually an expression of the *completeness* of the set of projection operators.

Suppose now that the $\mathbf{u}_k$ are eigenvectors of a bounded Hermitian operator[13].

$$\hat{A}\mathbf{u}_k = \alpha_k \mathbf{u}_k.$$

Then,

$$\hat{A}\mathbf{X}) = \sum_j A\mathbf{u}_j)(\mathbf{u}_j, \mathbf{X})$$

$$= \sum_j \alpha_j \hat{\pi}_j \mathbf{X}).$$

Thus,

$$\hat{A} = \sum_{j=1}^{N} \alpha_j \hat{\pi}_j. \tag{2.77}$$

This equation represents the *spectral decomposition* of the operator $\hat{A}$. If $\mathbf{X}$ and $\mathbf{Y}$ are arbitrary vectors, we have,

$$(\mathbf{Y}, \mathbf{X}) = \sum_j (\mathbf{Y}, \hat{\pi}_j \mathbf{X}), \tag{2.78a}$$

$$(\mathbf{Y}, \hat{A}\mathbf{X}) = \sum_j \alpha_j (\mathbf{Y}, \hat{\pi}_j \mathbf{X}). \tag{2.78b}$$

Eq. (2.77) is applicable when the eigenvalues of $\hat{A}$ are discrete. When the eigenvalues are continuous, it is possible to define a family of projection operator $\hat{\pi}_\alpha$, where α is real and continuous, with the following properties (see, Ref. 1, Sections 13 and 14).

(i) If $\alpha \le \beta$, then $\hat{\pi}_\alpha \le \hat{\pi}_\beta$, or $\hat{\pi}_\alpha \hat{\pi}_\beta = \hat{\pi}_\alpha = \hat{\pi}_\beta \hat{\pi}_\alpha$

(ii) For $a \le \alpha \le b$ and for $\epsilon > 0, \hat{\pi}_{\alpha+\epsilon}\psi \to \hat{\pi}_\alpha \psi$ as $\epsilon \to 0$, for any vector ψ.

(iii) $\hat{\pi}_\alpha \psi \to 0$ as $\alpha \to a$, and $\hat{\pi}_\alpha \psi \to \psi$ as $\alpha \to b$. That is $\hat{\pi}_\alpha = \hat{0}$ for $\alpha \le a$, and $\hat{\pi}_\alpha = \hat{1}$ for $\alpha \ge b$. and $\hat{\pi}_\alpha = \hat{1}$ for $\alpha \ge b$.

Corresponding to Eqs. (2.76), (2.77), (2.78a) and (2.78b), we have,

$$\int_a^b d\hat{\pi}_\alpha = \hat{1}, \text{ where } d\hat{\pi}_\alpha = \hat{\pi}_\alpha - \hat{\pi}_{\alpha-\epsilon}, \tag{2.79a}$$

$$\hat{A} = \int_a^b f(\alpha) d\hat{\pi}_\alpha \tag{2.79b}$$

$$(\phi, \psi) = \int_a^b d(\phi, \hat{\pi}_\alpha \psi), \tag{2.79c}$$

$$(\phi, \hat{A}\psi) = \int_a^b f(\alpha) d(\phi, \hat{\pi}_\alpha \psi). \tag{2.79d}$$

$d\hat{\pi}_\alpha$ is called a *differential projection operator.*

---

13. Remember that the eigenvectors of a bounded Hermitian operator form a complete set.

In Eq. (2.79b), $f(\alpha)$ is a function of α, which depends on the nature of the operator $\hat{A}$:

If $\hat{A}$ is Hermitian : $f(\alpha) = \alpha$; $a = b = -\infty$ (2.80a)

If $\hat{A}$ is Unitary : $f(\alpha) = e^{i\alpha}$; $a = 0$; $b = 2\pi$. (2.80b)

(Recall that the eigenvalues of a Unitary operator are unimodular).

The spectral decomposition (2.79b) is valid even in the case of Hermitian and Unitary operators which have no eigenvalues, as in the case of operators on infinite-dimensional spaces[14]. In this case, α would be merely a parameter. (In the case of operators with eigenvalues, α is related to the eigenvalues).

Problem 2.12: Show that, if $\hat{\pi}_\alpha \leq \hat{\pi}_\beta$, then

$$\hat{\pi}_\alpha \hat{\pi}_\beta = \hat{\pi}_\alpha = \hat{\pi}_\beta \hat{\pi}_\alpha$$

Problem 2.13: If $\hat{\pi}_1, \hat{\pi}_2, \ldots \hat{\pi}_N$ are projection operators, show that the sum of these operators are also projection operators if, and only if,

$$\sum_{k=1}^{N} (\mathbf{X}, \hat{\pi}_k \mathbf{X}) \leq (\mathbf{X}, \mathbf{X}),$$

for any vector $\mathbf{X}$ in the Hilbert space.

Problem 2.14: Consider the operator $\hat{p} \equiv -i\hbar \dfrac{d}{dx}$ in space $L^2(-\infty, \infty)$. The eigenvectors of $\hat{p}$ are $\mathbf{u}_k(x) = \dfrac{1}{\sqrt{2\pi}} e^{ikx}$, where k is continuous; $-\infty \leq k \leq \infty$. The eigenvalues are given by $\hat{p}\mathbf{u}_k(x) = \hbar k \mathbf{u}_k(x)$. Show that the differential projection operators $d\hat{\pi}_k$ are given by

$$d\hat{\pi}_k = \mathbf{u}_k) dk (\mathbf{u}_k.$$

Commuting Operators

We have mentioned that operators do not, in general, commute. However, there are operators which do commute. Such operators are of importance in quantum mechanics because the basis vectors in the Hilbert space of a physical system (As we will see in the next chapter, these basis vectors represent the quantum 'states' of the system) are determined by these. In this connection, the following theorem concerning two Hermitian operators, is basic:

14. Whereas every Hermitian or Unitary operator on a finite-dimensional space should have at least one eigenvalue, such operators on an infinite-dimensional space may or may not have eigenvalues.

Theorem: *If two bounded Hermitian operators commute, then they possess a complete orthonormal set of common eigenvectors, and vice versa.*

Now, it was stated earlier that the eigenvectors of a bounded Hermitian operator form a complete set. Therefore, we need show only that, if $\hat{A}$ and $\hat{B}$ are Hermitian operators such that $\hat{A}\hat{B} = \hat{B}\hat{A}$ then they have common eigenvectors. Conversely, if $\hat{A}$ and $\hat{B}$ have common eigenvectors, then $\hat{A}\hat{B} = \hat{B}\hat{A}$.

Let $$\hat{A}\mathbf{X} = \alpha\mathbf{X}. \tag{2.81}$$

We have to show that $$\hat{B}\mathbf{X} = \beta\mathbf{X}. \tag{2.82}$$

Multiplying Eq. (2.81) from the left by $\hat{B}$, we get,

$$\hat{B}(\hat{A}\mathbf{X}) = \hat{A}(\hat{B}\mathbf{X}) = \alpha(\hat{B}\mathbf{X}),$$

since $\hat{B}\hat{A} = \hat{A}\hat{B}$.

Thus, $\hat{B}\mathbf{X}$ is an eigenvector of $\hat{A}$ belonging to eigenvalue α.

If α is non-degenerate, then, $\hat{B}\mathbf{X}$ should be linearly dependent on $\mathbf{X}$, so that,

$$a(\hat{B}\mathbf{X}) + b\mathbf{X} = \mathbf{0}, \text{ with } a \neq 0; b \neq 0.$$

Or, $$\hat{B}\mathbf{X} = -(b/a)\mathbf{X} = \beta\mathbf{X}.$$

If α is g-fold degenerate, then, there are g linearly independent vectors $\mathbf{X}_k$ ($k = 1, 2, \ldots g$) such that

$$\hat{A}\mathbf{X}_k = \alpha\mathbf{X}_k.$$

We will assume that these are orthonormal:

$$(\mathbf{X}_k, \mathbf{X}_j) = \delta_{kj}. \tag{2.83}$$

We further assume, for the sake of simplicity, that $g = 2$. Let us define,

$$\mathbf{X} = c_1\mathbf{X}_1 + c_2\mathbf{X}_2 \tag{2.84}$$

where c_1 and c_2 are scalars. Since $\hat{A}$ is linear [see, Eq. (2.41a)].

$$\hat{A}\mathbf{X} = c_1(\hat{A}\mathbf{X}_1) + c_2(\hat{A}\mathbf{X}_2) = \alpha\mathbf{X}.$$

That is, $\mathbf{X}$ is an eigenvector of $\hat{A}$ belonging to eigenvalue α. What we have to prove now is reduced to showing that there are nonzero scalars c_1 and c_2 such that $\mathbf{X}$ defined by Eq. (2.84), satisfies Eq. (2.82):

$$\hat{B}\mathbf{X} \equiv \hat{B}(c_1\mathbf{X}_1 + c_2\mathbf{X}_2) = \beta(c_1\mathbf{X}_1 + c_2\mathbf{X}_2). \tag{2.85}$$

Taking the scalar product of Eq. (2.85) successively with $\mathbf{X}_1$ and $\mathbf{X}_2$, and using Eq. (2.83), we get,

$$(B_{11} - \beta)c_1 + B_{12}c_2 = 0,$$

$$B_{21}c_1 + (B_{22} - \beta)c_2 = 0. \tag{2.86}$$

where, we have used the abbreviation,

$$B_{jk} \equiv (\mathbf{X}_j, \hat{B}\mathbf{X}_k). \tag{2.87}$$

(2.86) is a set of homogeneous equations in c_1 and c_2. The condition for the existence of a nontrivial solution ($c_1, c_2 \neq 0$) is that the determinant of the coefficient matrix be zero, (see Section A.6). That is,

$$\begin{vmatrix} (B_{11}-\beta) & B_{12} \\ B_{21} & (B_{22}-\beta) \end{vmatrix} = 0. \tag{2.88a}$$

or $$B^2-(B_{11}+B_{22})\beta+(B_{11}B_{22}-|B_{12}|^2)=0, \tag{2.88b}$$

since $$B_{21} \equiv (\mathbf{X}_2, \hat{B}\mathbf{X}_1) = (\hat{B}\mathbf{X}_1, \mathbf{X}_2)^* = (\mathbf{X}_1, \hat{B}\mathbf{X}_2)^* = B_{12}^*.$$

The two roots of Eq. (2.88b) are,

$$\beta_1 = \frac{B_{11}+B_{22}}{2} + \frac{1}{2}\{(B_{11}+B_{22})^2 - 4(B_{11}B_{22}-|B_{12}|^2)\}^{1/2}, \tag{2.89a}$$

and $$\beta_2 = \frac{B_{11}+B_{22}}{2} - \frac{1}{2}\{(B_{11}+B_{22})^2 - 4(B_{11}B_{22}-|B_{12}|^2)\}^{1/2}. \tag{2.89b}$$

The two roots are equal, if and only if,

$$(B_{11}+B_{22})^2 - 4(B_{11}B_{22}-|B_{12}|^2)=0.$$

That is, if $$(B_{11}-B_{22})^2+4|B_{12}|^2=0. \tag{2.90}$$

Since both the terms here are real and positive (Problem 2.8), Eq. (2.90) will be satisfied only when each of the terms is zero. Thus, $\beta_1 = \beta_2$ only if

$$B_{11}=B_{22} \text{ and } B_{12}=B_{21}=0. \tag{2.91}$$

In this case,

$$\beta_1=\beta_2=\beta=B_{11}=B_{22}. \tag{2.92}$$

From Eqs. (2.91) and (2.92), we get (with the definition (2.87)),

$$\hat{B}\mathbf{X}_1=\beta\mathbf{X}_1;\ \hat{B}\mathbf{X}_2=\beta\mathbf{X}_2. \tag{2.93}$$

Thus, $\mathbf{X}_1$ and $\mathbf{X}_2$ are degenerate eigenvectors of $\hat{B}$ also. Eqs. (2.86) are satisfied for arbitrary values of c_1 and c_2 since the coefficients $(B_{11}-\beta)$, $(B_{22}-\beta)$, B_{12} and B_{21} are all zero, so that any linear combination of $\mathbf{X}_1$ and $\mathbf{X}_2$ are simultaneous (common) eigenvectors of $\hat{A}$ and $\hat{B}$.

If $\beta_1 \neq \beta_2$, then, one or both of the following conditions are satisfied:

$$(i)\ B_{11} \neq B_{22},\ \ (ii) B_{12} \neq 0.$$

If only (i) is satisfied, we get from Eqs. (2.89),

$$\beta_1 = B_{11};\ \beta_2 = B_{22}. \tag{2.94}$$

Corresponding to these values of β, we get two sets of values for c_1 and c_2. Denoting the values of c_1 and c_2 corresponding to β_k by $c_1^{(k)}$ and $c_2^{(k)}$, we have[15],

$$c_1^{(1)}=1,\ c_2^{(1)}=0;$$

$$c_1^{(2)}=0,\ c_2^{(2)}=1,$$

15. We assume $\mathbf{X}$ is normalized, so that, $|c_1^{(k)}|^2+|c_2^{(k)}|^2=1$.

so that,

$$\mathbf{X}^{(1)} = c_1^{(1)}\mathbf{X}_1 + c_2^{(1)}\mathbf{X}_2 = \mathbf{X}_1, \tag{2.95a}$$

$$\mathbf{X}^{(2)} = \mathbf{X}_2. \tag{2.95b}$$

Thus, $$\hat{B}\mathbf{X}_1 = \beta_1\mathbf{X}_1;\ \hat{B}\mathbf{X}_2 = \beta_2\mathbf{X}_2. \tag{2.96}$$

That is, $\mathbf{X}_1$ and $\mathbf{X}_2$ are eigenvectors of $\hat{B}$ belonging to different eigenvalues. When $B_{12} \neq 0$, we will similarly get,

$$\mathbf{X}^{(1)} = c_1^{(1)}\mathbf{X}_1 + c_2^{(1)}\mathbf{X}_2,$$

$$\mathbf{X}^{(2)} = c_1^{(2)}\mathbf{X}_1 + c_2^{(2)}\mathbf{X}_2,$$

with $c_1^{(k)} \neq 0$ and $c_2^{(k)} \neq 0$, such that,

$$\hat{B}\mathbf{X}^{(k)} = \beta_k\mathbf{X}^{(k)},\ k = 1, 2. \tag{2.97}$$

We can summarise the different possibilities:

(1) $\hat{A}$ has no degenerate eigenvalues. In this case, every eigenvector of $\hat{A}$ is also an eigenvector of $\hat{B}$.

(2) $\hat{A}$ has degenerate eigenvalues. The following possibilities are there:

  (a) Every eigenvector of $\hat{A}$ is also an eigenvector of $\hat{B}$

    (i) The degenerate eigenvectors[16] of $\hat{A}$ are degenerate eigenvectors of $\hat{B}$ also.

    (ii) The degenerate eigenvectors of $\hat{A}$ belong to different eigenvalues of $\hat{B}$. In this case, we say that the degeneracy is *lifted* (or, removed) by the Hermitian operator $\hat{B}$.

  (b) Every degenerate eigenvector of $\hat{A}$ is not an eigenvector of $\hat{B}$. But there are linear combinations of the degenerate eigenvectors, as many in number as the degree of degeneracy, which are degenerate eigenvectors of $\hat{A}$ but are non-degenerate eigenvectors of $\hat{B}$. The degeneracy is lifted by $\hat{B}$ in this case.

The foregoing conclusions, which were based on the results for $g = 2$, could be generalized to the case where $g > 2$. We note that Eq. (2.88a) is the secular equation of the matrix.

$$B \equiv \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{2.98}$$

Therefore, in the general case, we will have a $g \times g$ matrix in place of this 2×2 one. If all the g roots of the corresponding secular equation are different, the degeneracy with regard to the eigenvalue of $\hat{A}$ is completely removed by $\hat{B}$, and the common eigenvectors of $\hat{A}$ and $\hat{B}$ are the eigenvectors of the matrix B. These

16. We mean, by this, the linearly independent eigenvectors which belong to a degenerate eigenvalue.

eigenvectors are uniquely specified by the eigenvalues of $\hat{A}$ and $\hat{B}$. Thus, $\mathbf{X}^{(k)}$, which is the eigenvector belonging to eigenvalue α of $\hat{A}$ and the eigenvalue β_k of $\hat{B}$, could be written as

$$\mathbf{X}^{(k)} \equiv \psi_{\alpha,\beta_k}.$$

If some or all of the eigenvalues of the matrix $\hat{B}$ are equal, then the degeneracy is removed partly or not at all, respectively, by $\hat{B}$. In any case, there is a set of common orthonormal, eigenvectors for $\hat{A}$ and $\hat{B}$.

As to the second part of the theorem, let $\{\mathbf{u}_k\}$ be a set of common eigenvectors for $\hat{A}$ and $\hat{B}$:

$$\hat{A}\mathbf{u}_k = \alpha_k \mathbf{u}_k;\ \hat{B}\mathbf{u}_k = \beta_k \mathbf{u}_k.$$

Then, $\hat{A}(\hat{B}\mathbf{u}_k) = \hat{A}(\beta_k \mathbf{u}_k) = \beta_k(\hat{A}\mathbf{u}_k) = \beta_k \alpha_k \mathbf{u}_k$, since $\hat{A}$ is linear.

Similarly, $\hat{B}(\hat{A}\mathbf{u}_k) = \hat{B}\alpha_k \mathbf{u}_k = \alpha_k \beta_k \mathbf{u}_k = \hat{A}\hat{B}\mathbf{u}_k$.

An arbitrary vector $\mathbf{X}$ in the space can be written as,

$$\mathbf{X} = \sum_k x_k \mathbf{u}_k,$$

so that,

$$\hat{A}\hat{B}\mathbf{X} = \hat{B}\hat{A}\mathbf{X},$$

and this, according to Eq. (2.44a), implies that $\hat{A}\hat{B} = \hat{B}\hat{A}$.

Illustrative Example:

Consider the operator $\hat{H} \equiv -\dfrac{\hbar^2}{2m}\dfrac{d^2}{dx^2}$, on space $L^2(-\infty, \infty)$. This is Hermitian since it is the square of the Hermitian operator $-\dfrac{i\hbar}{\sqrt{2m}}\left(\dfrac{d}{dx}\right)$ (see, Problem 2.9).

The eigenvectors of $\hat{H}$ are (with $p \geq 0$)

$$\psi_1 = \frac{1}{\sqrt{2\pi}} \exp\left[-(i/\hbar)px\right] \text{ and } \psi_2(x) = \frac{1}{\sqrt{2\pi}} \exp\left[(i/\hbar)px\right],$$

both of which belong to the same eigenvalue $E = \dfrac{p^2}{2m}$:

$$\hat{H}\psi_1(x) = \left(\frac{p^2}{2m}\right)\psi_1(x);\ \hat{H}\psi_2(x) = \left(\frac{p^2}{2m}\right)\psi_2(x).$$

The operator $\hat{p} \equiv -i\hbar\dfrac{d}{dx}$ is Hermitian and commutes with $\hat{H}$ since any scalar operator commutes with itself and, hence, with its square. Vector operators, however, need not have this property (see Section 5.1). Now,

$$\hat{p}\psi_1(x) = -i\hbar\frac{d}{dx}\left(\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{i}{\hbar}px\right]\right) = -p\psi_1(x)$$

$$\hat{p}\psi_2(x) = p\psi_2(x), \text{ so that,}$$

$$\psi_1 \equiv \psi_{-p}; \psi_2 \equiv \psi_p.$$

Thus, ψ_1 and ψ_2 are non-degenerate eigenvectors of $\hat{p}$. The degeneracy with regard to the eigenvalues of $\hat{H}$, is lifted by $\hat{p}$.

Note: As we will see in the next chapter [Eqs. (3.18[1]) and (3.9[2])] $\hat{H}$ and $\hat{p}$ respectively correspond to the energy and the momentum of a free particle. ψ_1 and ψ_2 represent particles with opposite momenta but with the same energy.

Complete Set of Commuting Operators

Consider a bounded Hermitian operator $\hat{A}$. Its eigenvectors form a complete, orthonormal set and, thus, constitute a basis in the linear vector space on which $\hat{A}$ is defined. If $\hat{A}$ has degenerate eigenvalues, the eigenvectors are not uniquely specified by the eigenvalues of $\hat{A}$. That is, the basis is not unique [see, Eq. 2.84: Instead of $\mathbf{X}_1$ and $\mathbf{X}_2$, we could choose $\mathbf{X}^{(1)} = c_1^{(1)}\mathbf{X}_1 + c_2^{(1)}\mathbf{X}_2$ and $\mathbf{X}^{(2)} = c_1^{(2)}\mathbf{X}_1 + c_2^{(2)}\mathbf{X}_2$, with $|c_1^{(k)}|^2 + |c_2^{(k)}|^2 = 1$ and $c_1^{(1)*}c_1^{(2)} + c_2^{(1)*}c_2^{(2)} = 0$. A possible choice is $c_1^{(1)} = c_2^{(2)} = \cos\theta$; $c_2^{(1)} = -c_1^{(2)} = \sin\theta$. All the four vectors $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, belong to the same eigenvalue α of $\hat{A}$]. We should, then, seek another Hermitian operator $\hat{B}$ that commutes with $\hat{A}$. If the common eigenvectors of $\hat{A}$ and $\hat{B}$ are now uniquely specified by the eigenvalues of $\hat{A}$ and $\hat{B}$, we say that $\hat{A}$ and $\hat{B}$ form a *complete set* of operators. If this is not the case, then, we should find a third operator $\hat{C}$ which commutes with $\hat{A}$ and $\hat{B}$, and so on, until we have a set $\hat{A}, \hat{B}, \hat{C}, \ldots, \hat{L}$ of mutually commuting operators such that they have one, and *only* one, common basis. In this case, a common eigenvector is uniquely specified by the eigenvalues $a, b, \ldots l$ of the operators:

$$\psi \equiv \psi_{a,b,c\ldots l},$$

where

$$\hat{A}\psi_{a,b,c,\ldots l} = a\psi_{a,b,c\ldots l}.$$

$$\hat{L}\psi_{a,b,c,\ldots l} = l\psi_{a,b,c,\ldots l}.$$

$\hat{A}, \hat{B}, \ldots \hat{L}$ are, then, said to form a complete set of commuting operators.

Any operator $\hat{Q}$ which commutes with each of the operators $\hat{A}$, $\hat{B}$, ..., $\hat{L}$, will have necessarily the common basis of $\hat{A}$, $\hat{B}$, ..., $\hat{L}$ as its basis and its eigenvalues q would be functions of $a, b, \ldots l$:

$$q \equiv q(a, b, \ldots l).$$

In other words, $\hat{Q}$ will be a function of the operators $\hat{A}, \hat{B}, \ldots \hat{L}$.

In our illustrative example, $\hat{p}$ by itself constitutes a complete set of operator for the one-dimensional system with no external field, whereas $\hat{H}$ by itself is not a complete set; in fact, $\hat{H}$ is a function of $\hat{p}$.

2.3 BRA AND KET NOTATION FOR VECTORS

The fact that the scalar product (X, Y) is linear in Y whereas it is anti-linear in X, suggests that it is advantageous to think of the two vectors as belonging to two different spaces linear in themselves, but related to each other antilinearly. This way, we can make the scalar product 'symmetrical' in X and Y.

Thus, we have a space of the prefactors and a space of the postfactors. A vector in the prefactor space is denoted by a *bra*, $<\ |$, whereas a vector in the postfactor space is denoted by a *ket*, $|\ >$. Thus, X and Y in (X, Y) are written, respectively, as $<X\ |$ and $|\ Y>$, and the scalar product of Y by X as[17] $<x\ |\ Y>$. That is,

$$(\mathbf{X},\ \mathbf{Y}) \equiv\ <X\ |\ Y>\ =\ <Y\ |\ X>^{*}. \tag{2.99}$$

The prefactor space is, thus, a *bra-space* and the postfactor space a *ket-space.* Since the conjugate of a product of complex functions (to which the vectors in the Hilbert space bear analogy) is the product of the conjugates, Eq. (2.99) implies that

$$|\ X>^{*}=<X\ | \tag{2.100}$$

This shows that the two spaces are not independent of each other; they are said to be *dual* to each other. Not only is there a vector in one space corresponding to every vector in the other space, but also each relationship among vectors in one space has its 'image' in the other space. Some of these relationships in the ket-space and their image in the bra space are listed below :

Ket Space	Bra Space
$\|X>$	$<X\|$
$c\ \|X>$	$<X\|c^{*}$
$\|Z>\ =\ \|X>+\|Y>$	$<Y\|+<X\|\ =\ <Z\|$
$\|Y>\ =\hat{A}\ \|X>$	$<X\|\overline{A}=\ <Y\|$

Note that, in the bra space, operators act from 'right to left'. A scalar c in the ket space becomes its complex conjugate in the bra space, whereas an operator $\hat{A}$ in the ket space is transformed into the operator $\overline{A}$ in the bra space. Since we have not defined what $\overline{A}$ is, its relationship to $\hat{A}$ is to be found out (that is, we have to determine whether $\overline{A}$ is the Hermitian conjugate (adjoint) of $\hat{A}$, the transpose[18] of $\hat{A}$, or some other way related to $\hat{A}$). For this, consider the scalar product, $<Z\ |\ Y>$, where $|\ \mathbf{Y}=\hat{A}\ |\ \mathbf{X}>$.

---

17. $<\ |\ >$ is, actually, a short form of $<||>$. Thus, a scalar product is nothing but the product of a bra vector and a ket vector in that order.

18. $\hat{B}$ is the transpose of $\hat{A}$ if for arbitrary X and Y in the common domain of $\hat{A}$ and $\hat{B}$, $(\mathbf{X}, \hat{A}\ \mathbf{Y}) = (\mathbf{Y}, \hat{B}\ \mathbf{X})$.

From $< Z \mid Y > = < Y \mid \mathbf{Z} >^*$, we have,

$$< Z \mid \{\hat{A} \mid X >\} = [\{< X \mid \overline{\hat{A}}\} \mid Z >]^* \tag{2.101}$$

But $$< Z \mid \{\hat{A} \mid X >\} \equiv (Z, \hat{A}\mathbf{X}) = (\hat{A}^\dagger Z, X) = (X, \hat{A}^\dagger Z)^*$$

$$\equiv [< X \mid \{\hat{A}^\dagger \mid Z >\}]^*.$$

Thus,

$$[\{< X \mid \overline{\hat{A}}\} \mid Z >]^* = [< X \mid \{\hat{A}^\dagger \mid Z >\}]^* \tag{2.102}$$

But from the definition (2.54) of the Hermitian conjugate, we know that $\hat{A}^\dagger$ operating on the post-factor in a scalar product is equivalent to $\hat{A}$ operating on the pre-factor. Therefore,

$$\overline{\hat{A}} = \hat{A}. \tag{2.103}$$

Also, from Eqs. (2.101) and (2.102), we have,

$$< Z \mid \hat{A} \mid X > \; = \; < X \mid \hat{A}^\dagger \mid Z >^*, \tag{2.104}$$

so that the condition for $\hat{A}$ to be Hermitian is

$$< Z \mid \hat{A} \mid X > \; = \; < X \mid \hat{A} \mid Z >. \tag{2.105}$$

Many of the relationships among vectors assume an elegant appearance when expressed in the new notation. Thus, we have,

Orthonormality (Eq. (2.16)) : $< u_i \mid u_j > \; = \delta_{ij}$ (2.16^1)

Completeness (Eq. (2.27)) : $\Sigma \mid u_j > \; < u_j \mid = \hat{1}.$ (2.27^1)

Projection operator (Eq. (2.74)) : $\hat{\pi}_j = \; \mid u_j > \; < u_j \mid .$ (2.74^1)

2.4 REPRESENTATION THEORY

Consider an orthonormal basis $[\mid u_j >]_N$ in an N-dimensional space. Any vector in the space can be expanded in terms of the vectors $\mid u_j >$. Thus, if $\mid X >$ and $\mid Y >$ are arbitrary vectors, we have,

$$\mid X > = \sum_{j=1}^{N} x_j \mid u_j >, \tag{2.106a}$$

$$\mid Y > = \sum_{j=1}^{N} y_j \mid u_j >, \tag{2.106b}$$

If $\mid Z >$ is a vector such that

$$|Z> = a\,|X> + b\,|Y>, \tag{2.106c}$$

then,

$$z_j = ax_j + by_j \tag{2.107}$$

where, $$|Z> = \sum_{j=1}^{N} z_j\,|u_j>.$$

Also, $$<X\,|\,Y> = \sum_{j=1}^{N} x_j^* y_j. \tag{2.108}$$

Eqs. (2.107) and (2.108) suggest that, in place of the abstract vectors $|X>, |Y>, \ldots$, we can deal with their *ordered* expansion coefficients (or, components), $[x] \equiv [x_1, x_2, \ldots, x_N], [y] \equiv [y_1, y_2, \ldots, y_N]$, etc. These ordered expansion coefficients will be called the *representatives* of the vectors. Corresponding to every relationship between vectors, there is a relationship between representatives. Thus, the relationship (2.106c) translates as,

$$[z] = a[x] + b[y], \tag{2.109a}$$

or $$[z_1, z_2, \ldots, z_N] = [ax_1 + by_1, ax_2 + by_2, \ldots, ax_N + by_N] \tag{2.109b}$$

Unlike the vectors, the representatives depend upon the basis chosen; changing the basis will change the representatives also. However, with respect to a given basis, the representative $[x]$ corresponding to the vector $|X>$ is unique. We say that the vector $|X>$ is *represented* by $[x]$ in the *representation* defined by the basis $[|u_j>]_N$. The basis vectors themselves are represented by $[u_1]$, $[u_2]$, ..., $[u_N]$.

where,
$$[u_1] \equiv [1, 0, 0, \ldots 0],$$
$$\vdots$$
$$[u_2] \equiv [0, 1, 0, \ldots 0],$$
$$\vdots$$
$$[u_N] \equiv [0, 0, \ldots\ldots 1].$$

Eqs. (2.109a, b) and (2. 108) show that the representative $[x]$ of the ket vector $|X>$ could be written in the form of a column matrix x:

$$|X> \;\rightarrow [x] \rightarrow x \equiv \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_N \end{pmatrix}, \tag{2.111a}$$

whereas, the representative of the bra vector $<X|$ is represented by the row matrix x^+ [see Eqs. (A.33) and (A.6)]:

$$< X \mid \rightarrow x^{\dagger} = (x_1^* x_2^* \ldots x_N^*) \tag{2.111b}$$

The vector-addition (2.106c) is, then, represented by the matrix addition:

$$z = ax + by,$$

and the scalar product (2.108) by the matrix product, $x^{\dagger}y$.

The unit vectors $\mid u_i >$ are represented by,

$$\mid u_i > \rightarrow u_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}; u_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}; \ldots u_N = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 1 \end{pmatrix} \tag{2.112}$$

The orthonormality condition (2.16[1]) and the completeness condition (2.27[1]) become,

$$< u_i \mid u_j > \rightarrow \mathbf{u}_i^{\dagger} \mathbf{u}_j = \delta_{ij} \tag{2.113a}$$

and

$$\sum_j \mid u_j > \; < u_j \mid \rightarrow \sum_j u_j u_j^{\dagger} = I \tag{2.113b}$$

The operator equation,

$$\hat{A} \mid X > \; = \; \mid Y >, \tag{2.114a}$$

is represented by the matrix equation,

$$Ax = y. \tag{2.114b}$$

Since both x and y are $(N \times 1)$ matrices, A has to be an $(N \times N)$ matrix. Thus, an operator in an N-dimensional space, is represented by a *square matrix* of order N. The properties of the linear operators, thus, follow from the properties of the square matrices. In particular, the operator algebra is identical with the matrix algebra of square matrices. The eigenvalue problem for an operator is reduced to the problem of diagonalizing a square matrix.

The above-described scheme of representing vectors and operators by matrices, is referred to as *Matrix representation.*

It follows that a Hermitian operator would be represented by a Hermitian matrix and a Unitary operator by a Unitary matrix :

For, according to the definition (2.103) of a Hermitian operator, we have

$$< X \mid \hat{A} \mid Y >^* = \; < Y \mid \hat{A} \mid X >,$$

i.e, in matrix representation reads,

$$(X^{\dagger} A Y)^{\dagger} = Y^{\dagger} A X.$$

$$Y^{\dagger} A^{\dagger} X = Y^{\dagger} A X, \text{ or, } A^{\dagger} = A.$$

Similarly, if $\hat{U}$ is Unitary, we have,

$$< X \mid \hat{U}^{\dagger}\hat{U} \mid Y > \; = \; < X \mid \hat{U}\hat{U}^{\dagger} \mid Y > \; = \; < X \mid Y >$$

i.e. $$X^{\dagger}U^{\dagger}UY = X^{\dagger}UU^{\dagger}Y = X^{\dagger}Y$$

or $$U^{\dagger}U = UU^{\dagger} = I.$$

Also, an operator on an infinite-dimensional linear vector space, would be represented by an infinite-dimensional matrix.

Matrix Elements of A

Consider the equation (2.114a)

$$\mid Y > \; = \hat{A} \mid X > .$$

Expanding $\mid Y >$ and $\mid X >$ in terms of the unit vectors $\mid u_i >$, we get,

$$\sum_{i=1}^{N} y_i \mid u_i > = \hat{A} \sum_{k=1}^{N} x_k \mid u_k >$$

$$= \sum_{k=1}^{N} x_k \hat{A} \mid u_k >,$$

since $\hat{A}$ is linear.

Taking the scalar product of the equation with $\mid u_j >$, we have, using Eq. (2.113a),

$$y_j = \sum_{k} < u_j \mid \hat{A} \mid u_k > x_k. \tag{2.115}$$

But, according to Eq. (2.114b), we have (see Eq. (A.7)),

$$y_j = A_{jk} x_k. \tag{2.116}$$

Comparing Eqs. (2.115) and (2.116), we see that

$$A_{jk} \equiv \; < u_j \mid \hat{A} \mid u_k > . \tag{2.117a}$$

That is, the jk–th matrix element of the matrix A that represents the operator $\hat{A}$ in the representation defined by the basis $[\mid u_j >]_N$, is the scalar product of the vector $\hat{A} \mid u_k >$ by the vector $\mid u_j >$.

If $\hat{A}$ is one of the complete set of commuting operators that define the basis (see the last part of Section 2.2), then,

$$\hat{A} \mid u_k > \; = \alpha_k \mid u_k >,$$

so that, $$A_{jk} = \delta_{jk} \alpha_k. \tag{2.117b}$$

Thus, A is diagonal. That is, an operator is represented by a diagonal matrix in a representation defined by its own eigenvectors. This result is consistent with the fact that the eigenvalues of a diagonal matrix are its diagonal elements (Eq. (A. 49)).

Change of Basis

Since the matrices (that is, their matrix elements) representing vectors and operators depend on the representation, or basis, we are faced with the problem of finding the relationship between matrices which represent the same set of vectors and operators in different representations. We address ourselves to a solution of this problem in the following :

Let $[|u_i>]_N$ and $[|u_i'>]_N$ be two orthonormal bases in a Hilbert space. Since both the sets are complete, the vectors of one set can be expanded in terms of the vectors in the other set :

$$|u_j'> = \sum_{k=1}^{N} |u_k> S_{kj}, j = 1, 2, \ldots, N. \tag{2.118}$$

The expansion coefficients S_{kj} could be regarded as the matrix elements of an $(N \times N)$ matrix S which represents the transformation from the representation $[|u_i>]$ to the representation $[|u_i'>]$. From Eq. (2.118), we have (taking the scalar product of the equation by $|u_i>$),

$$S_{ij} = <u_i|u_j'> \tag{2.119}$$

That is, S_{ij} is the 'component' of $|u_j'>$ along $|u_i>$.

From the orthonormality of the set $[|u_i'>]$, we have,

$$\delta_{jk} = <u_k'|u_j'> = \sum_{i=1}^{N} <u_k'|u_i> <u_i|u_j'>$$

$$= \sum_{i=1}^{N} S_{ik}^* S_{ij} = (S^\dagger S)_{kj}, \tag{2.120a}$$

where, we have used the closure property of the basis $[|u_i>]_N$ (see Eq. (2.27[1]), and Eqs. (A.) and (A.33a)). Similarly, from the completeness property of the set $[|u_i'>]$, we have,

$$\delta_{jk} = <u_j|u_k> = \sum_{i=1}^{N} <u_j|u_i'> <u_i'|u_k>$$

$$= \sum_{i=1}^{N} S_{ji} S_{ki}^* = (SS^\dagger)_{jk}, \tag{2.120b}$$

where, the orthonormality of the set $[|u_i>]$ has been used. From Eqs. (2.120a, b) we have,

$$S^\dagger S = I = SS^\dagger, \tag{2.120}$$

Given an orthonormal basis, the first part of Eq. (2.120) represents the orthonormality while the second part the completeness, of the transformed basis. Change of (orthonormal) basis in a linear vector space is, thus represented by a Unitary matrix.

Eq. (2.118) can be written as a matrix equation if we define a matrix U by

$$U \equiv (u_1 u_2 \ldots u_N), \tag{2.121}$$

where, u_k is the column matrix representing the basis vector. $|u_k>$ (see Eq. (2.112)). Thus U is an $(N \times N)$ matrix[19]. The orthonormality of the basis (Eq. (2.113a)) requires that

$$U^\dagger U \equiv \begin{pmatrix} u_1^\dagger \\ u_2^\dagger \\ \cdot \\ \cdot \\ \cdot \\ u_N^\dagger \end{pmatrix} (u_1 u_2 \ldots u_N) = \begin{pmatrix} u_1^\dagger u_1 & u_1^\dagger u_2 & \ldots & u_1^\dagger u_N \\ u_2^\dagger u_1 & u_2^\dagger u_2 & \ldots & u_2^\dagger u_N \\ \ldots & \ldots & \ldots & \ldots \\ u_N^\dagger u_1 & u_N^\dagger u_2 & \ldots & u_N^\dagger u_N \end{pmatrix} = I, \tag{2.122a}$$

whereas, completeness (Eq. (2.113b)) requires that

$$UU^\dagger = \left(\sum_{i=1}^{N} u_i u_i^\dagger \right) = I, \tag{2.122b}$$

where I is the $(N \times N)$ unit matrix.

Thus, U is Unitary. That is, *an orthonormal basis can be represented by a Unitary matrix. Conversely, the columns and rows of a Unitary matrix represent orthonormal vectors.*

In terms of the matrices U and S, Eq. (2.118) reduces to

$$U' = US. \tag{2.118a}$$

Since both U and S are unitary we have also,

$$U = U'S^\dagger, \tag{2.118b}$$

and

$$S = U^\dagger U'. \tag{2.119a}$$

Eq. (2.118b) is the inverse of Eq. (2.118), whereas Eq. (2.119a) is the matrix equivalent of Eq. (2.119).

The expansion of an arbitrary vector $|X>$ in terms of the basis vectors is,

$$|X> = \sum_{i=1}^{N} x_i |u_i> = \sum_{j=1}^{N} x_j' |u_j'>, \tag{2.123a}$$

which shows that the product Ux is invariant:

$$Ux = U'x', \tag{2.123b}$$

where x and x' are the column matrices representing $|X>$ in the representation U and U', respectively.

Hence

$$x' = (U')^\dagger Ux = S^\dagger x, \tag{2.124a}$$

or,

$$x = Sx'. \tag{2.124b}$$

The linear transformation

$$|Y> = \hat{A} |X>,$$

is represented by the matrix equation,

---

19. U could also be regarded as a row-matrix whose elements are column matrices.

$$y = Ax, \tag{2.125a}$$

in the representation U and by the equation,

$$y' = A'x', \tag{2.125b}$$

in the representation U'. Substituting for y and x in (2.125a) from Eq. (2.124b), we get,

$$Sy' = ASx'$$

or,

$$y' = (S^{\dagger}AS)x'. \tag{2.125c}$$

Comparing Eqs. (2.125b) and (2.125c), we have,

$$A' = S^{\dagger}AS. \tag{2.126}$$

Eqs. (2.124a) and (2.126) represent, respectively, the transformation law for vectors and operators under change of basis[20]. We see that the *change of basis corresponds to a Unitary transformation.*

From Eqs. (2.119a) and (2.126), we see that,

$$U'A'U'^{\dagger} = UAU^{\dagger}, \tag{2.127}$$

which shows that the product $(UAU^{\dagger})$ is invariant under change of basis.

Eqs. (2.118a) and (2.124a) could be interpreted to mean the following; A 'rotation' of the basis vectors (the co-ordinate system) is equivalent to an inverse rotation of the physical vectors (see Section 5.6).

Problem 2.15 : An operator $\hat{A}$ is represented by the matrix $A_u = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ in a representation in which the basis vectors are $\mathbf{u}_1$ and $\mathbf{u}_2$. Obtain the matrix A_v that represents $\hat{A}$ in the representation in which the basis vectors are $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(\mathbf{u}_1 + \mathbf{u}_2)$ and $\mathbf{v}_2 = \frac{1}{\sqrt{2}}(\mathbf{u}_1 - \mathbf{u}_2)$.

---

20. The change of basis could also be described in terms of the Unitary matrix ; $T = S^{\dagger}$. In terms of T, we have,

$$x' = Tx, \tag{2.124b'}$$

$$A = TAT^{\dagger} \tag{2.126'}$$

The elements of T are given by $T_{ij} = S^{*}_{ji} = \langle u_i' \mid u_j \rangle$.

2.4A Co-ordinate and Momentum Representations

Often, we have to represent the vectors as functions of either the co-ordinates or the momenta. For example, the vectors in the function space $L^2(a,b)$ are such functions (see Section 2.1). We will here discuss the relationship between this function-space notation and the bra-ket notation.

We will denote by $<\mathbf{r}\mid u>$ the co-ordinate representation of the ket $\mid u>$. Similarly, $<\mathbf{p}\mid u>$ is the momentum representation of $\mid u>$. Thus,

$$\mathbf{u}_j(\mathbf{r}) = <\mathbf{r}\mid u_j>, \tag{2.128a}$$

$$\mathbf{u}_j(\mathbf{p}) = <\mathbf{p}\mid u_j>, \tag{2.129a}$$

Also,

$$\mathbf{u}_j^*(\mathbf{r}) = <u_j\mid \mathbf{r}>, \tag{2.128b}$$

$$\mathbf{u}_j^*(\mathbf{p}) = <u_j\mid \mathbf{p}>. \tag{2.129b}$$

The scalar product $<u_i\mid u_j>$ is given by

$$<u_i\mid u_j> \equiv (u_i,\mathbf{u}_j) = \int \mathbf{u}_i^*(\mathbf{r})\mathbf{u}_j(\mathbf{r})d^3\mathbf{r}$$

$$= \int <u_i\mid \mathbf{r}> \; <\mathbf{r}\mid u_j> d^3\mathbf{r} \tag{2.130}$$

Hence,

$$\int \mid \mathbf{r}> \; <\mathbf{r}\mid d^3\mathbf{r} = \hat{1}. \tag{2.131a}$$

Similarly,

$$\int \mid \mathbf{p}> \; <\mathbf{p}\mid d^3\mathbf{p} = \hat{1}. \tag{2.131b}$$

From Eqs. (2.129a) and (2.131a), we have,

$$\mathbf{u}_j(\mathbf{p}) \equiv <\mathbf{p}\mid u_j> = \int <\mathbf{p}\mid \mathbf{r}> \; <\mathbf{r}\mid u_j> d^3\mathbf{r} = \int \mathbf{u}_j(\mathbf{r}) <\mathbf{p}\mid \mathbf{r}> d^3\mathbf{r}, \tag{2.132}$$

and from (2.128a) and (2.131b),

$$\mathbf{u}_j(\mathbf{r}) = \int \mathbf{u}_j(\mathbf{p}) <\mathbf{r}\mid \mathbf{p}> d^3\mathbf{p} = \int \mathbf{u}_j(\mathbf{p}) <\mathbf{p}\mid \mathbf{r}>^* d^3\mathbf{p}. \tag{2.133}$$

Substituting p_x (x-component of $\mathbf{p}$) for $\mathbf{p}$ and x for $\mathbf{r}$ in (2.132), we have,

$$\mathbf{u}_j(p_x) = \int_{-\infty}^{+\infty} \mathbf{u}_j(x) <p_x\mid x> dx. \tag{2.132a}$$

Similarly, from (2.133), we get,

$$\mathbf{u}_j(x) = \int_{-\infty}^{+\infty} \mathbf{u}_j(p_x) <p_x\mid x>^* dp_x. \tag{2.133a}$$

Now, it will be suggested in Section 3.1 that dynamical variables are represented in quantum mechanics, by Hermitian operators (Postulate I). It is further shown there that the operators corresponding to the dynamical variables x and p_x are given by [Eqs. (3.18) and (3.18^1)],

$$\hat{x} = x;\ \hat{p}_x = -i\hbar\frac{d}{dx}\text{: (Co-ordinate representation),}$$

and $$\hat{x} = i\hbar\frac{d}{dp_x};\ \hat{p}_x = p_x\text{: (Momentum representation)}$$

Thus,
$$\hat{x}\mathbf{u}_j(x) = x\mathbf{u}_j(x), \tag{2.134a}$$

$$\hat{x}\mathbf{u}_j(p_x) = i\hbar\frac{d\mathbf{u}_j(p_x)}{dp_x}, \tag{2.134b}$$

$$\hat{p}_x\mathbf{u}_j(x) = -i\hbar\frac{d\mathbf{u}_j(x)}{dx}, \tag{2.135a}$$

$$\hat{p}_x\mathbf{u}_j(p_x) = p_x\mathbf{u}_j(p_x). \tag{2.135b}$$

From Eqs. (2.132a) and (2.134a), we have,

$$\hat{x}\mathbf{u}_j(p_x) = \int_{-\infty}^{+\infty}\{\hat{x}\mathbf{u}_j(x)\} < p_x \mid x > dx = \int_{-\infty}^{+\infty} x\mathbf{u}_j(x) < p_x \mid x > dx, \tag{2.136a}$$

while, from Eqs. (2.132a) and (2.134b), we get,

$$\hat{x}\mathbf{u}_j(p_x) \equiv i\hbar\frac{d\mathbf{u}_j(p_x)}{dp_x} = i\hbar\int_{-\infty}^{+\infty}\mathbf{u}_j(x)\left\{\frac{d}{dp_x} < p_x \mid x >\right\}dx \tag{2.136b}$$

Eqs. (2.136a, b) require that

$$\frac{d}{dp_x} < p_x \mid x > \ = -(i/\hbar)x < p_x \mid x >, \tag{2.137a}$$

or,
$$< p_x \mid x > \ = C_1 \exp\left[-(i/\hbar)p_x x\right], \tag{2.138a}$$

where C_1 is independent of p_x.

Similarly, from Eqs. (2.133a), (2.135b) and (2.135a), we get,

$$\frac{d}{dx} < p_x \mid x >^* = (i/\hbar)p_x < p_x \mid x >^*, \tag{2.137b}$$

so that
$$< p_x \mid x >^* = C_2^* \exp\left[(i/\hbar)p_x x\right],$$

or,

$$< p_x \mid x > = C_2 \exp\left[-(i/\hbar)p_x x\right], \tag{2.138b}$$

where, C_2 is independent of x.

Eqs. (2.138a, b) require that

$$C_1 = C_2 = C,$$

where, C is a constant (independent of both x and p_x). Thus,

$$< p_x \mid x >= C \exp\left[-(i/\hbar)p_x x\right] \tag{2.138}$$

Substituting (2.138) in (2.132a) and (2.133a), we get,

$$\mathbf{u}_j(p_x) = C \int_{-\infty}^{+\infty} \mathbf{u}_j(x) \exp\,[(-i/\hbar)p_x x]\,dx \tag{2.132b}$$

$$\mathbf{u}_j(x) = C^* \int_{-\infty}^{+\infty} \mathbf{u}_j(p_x) \exp\,[(i/\hbar)p_x x]\,dp_x \tag{2.133b}$$

The constant C can be determined from the (normalization) condition,

$$\int_{-\infty}^{+\infty} |\,u_j(x)\,|^2\,dx = 1$$

$$= |\,C\,|^2 \int_{-\infty}^{+\infty} dp_x \int_{-\infty}^{+\infty} dp'_x u_j^*(p'_x) u_j(p_x) \int_{-\infty}^{+\infty} \exp\,[(i/\hbar)(p_x - p'_x)x]\,dx$$

$$= 2\pi\hbar\,|\,C\,|^2,$$

Since

$$\int_{-\infty}^{+\infty} \exp\,[i/\hbar(p_x - p'_x)x]\,dx = 2\pi\hbar\delta(p_x - p'_x),$$

(see Eq. (D.6a))

and

$$\int_{-\infty}^{+\infty} |\,\mathbf{u}_j(p_x)\,|^2\,dp_x = 1.$$

Thus, assuming C to be real, we get,

$$C = C^* = \frac{1}{\sqrt{2\pi\hbar}}. \tag{2.139}$$

Eqs. (2.137a, b) suggest that, for the general case where ∇_p and ∇_r, respectively, take the place of $\dfrac{d}{dp_x}$ and $\dfrac{d}{dx}$, we should have,

$$\mathbf{u}_j(\mathbf{p}) = (2\pi\hbar)^{-3/2} \int \mathbf{u}_j(\mathbf{r}) \exp\,[(-i/\hbar)\mathbf{p}\cdot\mathbf{r}]\,d^3\mathbf{r} \tag{2.140a}$$

and

$$\mathbf{u}_j(\mathbf{r}) = (2\pi\hbar)^{-3/2} \int \mathbf{u}_j(\mathbf{p}) \exp\,[(i/\hbar)\mathbf{p}\cdot\mathbf{r}]\,d^3\mathbf{p} \tag{2.140b}$$

Thus, $\mathbf{u}_j(\mathbf{p})$ and $\mathbf{u}_j(\mathbf{r})$ are Fourier transforms of each other (see Eqs. (C.18a, b)).

Expression (2.117a) for the matrix elements of the operator $\hat{A}$ becomes:

$$A_{jk} = < u_j \mid \hat{A} \mid u_k > = \int\int < u_j \mid \mathbf{r} > \; < \mathbf{r} \mid \hat{A} \mid \mathbf{r}' > \; < \mathbf{r}' \mid u_k > d^3\mathbf{r} d^3\mathbf{r}'$$

$$= \int\int u_j^*(\mathbf{r}) \hat{A}(\mathbf{r}, \mathbf{r}') u_k(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}', \tag{2.141}$$

where, use has been made of Eqs. (2.128a, b) and (2.131a).

Here, $\hat{A}(\mathbf{r}, \mathbf{r}') \equiv < \mathbf{r} \mid \hat{A} \mid \mathbf{r}' >$ is the co-ordinate representation of $\hat{A}$, and is generally of the form,

$$\hat{A}(\mathbf{r},\mathbf{r}') = \hat{A}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}'). \tag{2.141a}$$

Thus,

$$A_{jk} = \int u_j^*(\mathbf{r})\hat{A}(\mathbf{r})u_k(\mathbf{r})d^3\mathbf{r}. \tag{2.117c}$$

Comparison of Eqs. (2.132) and (A.85c) shows that the former is merely Eq. (2.118) for the case where the transformation is represented by a continuous matrix. The other equations of this Section could also be similarly interpreted in terms of continuous matrices.

REFERENCES

1. Thomas F. Jordan, 1969. *Linear Operators for Quantum Mechanics*. (John Wiley, New York), Chapters I to IV.
2. G. L. Trigg, 1968. *Quantum Mechanics*. (Van Nostrand, East-West Edition), Chapter 3.

CHAPTER 3

THE BASIC PRINCIPLES

3.1 THE FUNDAMENTAL POSTULATES

Certain hypotheses, or postulates, serve as the foundation on which a physical theory is built up. The theory considers these postulates as fundamental. In other words, an *explanation* of the postulates is beyond the scope of the theory; the theory is rather concerned with the *consequences* of the postulates. In fact, the theory is nothing but the mathematical framework which connects the postulates to their observable consequences. Of course, one often regards the postulates as part of the theory itself, in which case the consequences are referred to as *predictions* of the theory. Agreement between these predictions and the experimental observations provides a *verification* of the theory, and, thus, an indirect justification of the postulates. This section is devoted to an enumeration and discussion of such experimentally justified postulates which embody the basic principles on which the mathematical edifice of quantum mechanics is built up.

In Section 1.2, we presented a set of postulates which, we emphasized, could be regarded as the quantum mechanical laws of motion, analogous to Newton's laws of motion in classical mechanics. A formalism of quantum mechanics which logically follow from those postulates, is the *path-integral formalism* (PIF) developed by Feynman[1]. We had also indicated in Section 1.2, that the probability amplitudes $\psi(\mathbf{r}, t)$ and the dynamical variables, such as the Hamiltonian H and the momentum $\mathbf{p}$, could be regarded, respectively as vectors and operators in a Hilbert Space. The formalism of quantum mechanics which incorporates this viewpoint could be called the *Hilbert-space formalism* (HSF). Whereas the PIF has got certain advantages over the HSF, especially when concerned with the extension of quantum mechanics to quantum field theory, for many of the practical applications of non-relativistic quantum mechanics, the HSF proves to be simpler and elegant enough. Also, the two earliest versions of quantum mechanics, namely, Schrödinger's *wave mechanics* and Heisenberg's *matrix mechanics*, emerge as special cases of the HSF. In this chapter, we present the postulates in a form (and language) that is appropriate to the Hilbert-space formalism of quantum mechanics.

---

1. See, the book by R.P. Feynman and A.R. Hibbs (Footnote 9, Chapter 1).

Postulate I: *Corresponding to every observable physical quantity, there exists a Hermitian operator. The only measurable values of a physical observable are the various eigenvalues of the corresponding operator.*

In classical mechanics, physical observables are represented by *dynamical variables*, such as energy, angular momentum, etc., which are functions of certain basic variables such as co-ordinates and momenta. Therefore, the postulate is equivalent to stating that *every dynamical variable is represented by a Hermitian operator*[2].

The Hermitian character of the operator ensures that the eigenvalues are real (Section 2.2), which property is necessary if the eigenvalues are to be identified with measured quantities. We further assume that the Hermitian operator representing an observable is bounded so that it has got a complete set of eigenvectors.

Since the eigenvalues of a Hermitian operator are not, in general, at least wholly, continuous, Postulate I contains the important feature of discreteness (that is, the fact that a dynamical variable such as angular momentum, can have only certain allowed values) associated with every microscopic system.

Now, operators have meaning only with reference to a set of vectors on which they operate. This fact leads us to the second postulate.

Postulate II: *To every physical system, there corresponds an abstract Hilbert space. A state of the system is represented by a vector in this space upon which the operators corresponding to the observables act. Also, any vector in the space represents a possible state of the system.*

Since an arbitrary vector in the space can be expressed as a linear superposition of a complete set of orthonormal vectors, this postulate implies the principle of superposition (Eq. (1.3)).

In classical mechanics, the dynamical state of a physical system is specified in terms of the values of a set of dynamical variables. Thus, giving the values of all the co-ordinates and momenta of the particles composing a system, at any instant, specifies the state of the system at that instant. This implies not only that the value of any other dynamical variable, such as energy, relevant to the system can be computed from the given values of the co-ordinates and momenta, but also that the state of the system at any other time can be deduced with the help of the equations of motion. The definition of the state of a system in quantum mechanics appears radically different from the above definition of the dynamical state in classical mechanics. However, the difference is only apparent. As we have already stated, any state vector can be expanded in terms of a complete set of basis vectors. These basis vectors are the common eigenvectors of a complete set of commuting Hermitian operators which, according to Postulate I, represent the dynamical variables of the system. In other words, we require the (possible)

2. Every Hermitian operator does not, however, represent an observable.

values of the dynamical variables of the system in order to specify the basis vectors in terms of which any state vector is defined. Thus, even though we can't speak of definite values for the dynamical variables in a given state, the concept of dynamical variables is required to specify a state. For example, the basis states for a spin-$\frac{1}{2}$ particle are the spin-up and the spin-down states, represented respectively by χ_+ and χ_-. A general state vector would be, then,

$$\chi = c_+\chi_+ + c_-\chi_- . \tag{3.1}$$

Thus, even though we cannot say whether the system has spin up or down in the state χ, we have to make use of the concept of the dynamical variable spin, and its possible values in order to define χ[3].

Now, the eigenvalues are obtained by operating on the state vector with the Hermitian operator (Eq.(2.49)). This, according to Postulate I, corresponds to the act of measurement. Thus, if we make a measurement of a certain physical observable represented by the operator $\hat{A}$, the system would be left in a state which is an eigenvector of $\hat{A}$. Then, if we make a measurement of another observable represented, say, by $\hat{B}$, this act of measurement will carry the system over to a state which is an eigenstate of $\hat{B}$. The two measurements will refer to the same state of the system only if the state is represented by a common eigenvector of $\hat{A}$ and $\hat{B}$. And this would be so (see commuting operators, Section 2.2) only if $\hat{A}$ and $\hat{B}$ commute. Thus, dynamical variables which could be simultaneously assigned definite values in a given state, are represented by commuting Hermitian operators. Such variables are called *compatible variables*. In contrast, variables which cannot be assigned specific values in a given state (represented by non-commuting operators), are called *complementary variables*. Energy and momentum are compatible variables for a free particle, whereas the x-co-ordinate and the x-component of the momentum are examples of complementary variables.

For a given system, there is a limit to the number of compatible variables. This limit is represented by the complete set of commuting Hermitian operators on the Hilbert space of the system. The simultaneous measurement of the set of compatible variables corresponds to the simultaneous determination of the eigenvalues of this complete set of operators. A *maximal* measurement of this sort provides the maximum amount of information that can be had on a quantum mechanical system. In fact, unique (or, complete) specification of a state of the system requires such a maximal measurement.

---

3. In view of Postulate I, the question arises whether the state χ is a physical state or not. That is, whether the particle can exist in the state χ or not. Since χ is not an eigenvector of the Hermitian operator corresponding to spin, we will not find the particle in the state χ if we make a measurement of spin but will find either in the state χ_+ or in the state χ_-. Thus, it would appear that only the basis states are realizable; that is, the particle is always in one or the other of the basis states and never in between. χ should, then, refer to the outcome of a large number of measurements on identical systems. That is, χ represents the state of an *ensemble* as far as the spin is concerned (see Postulate III).

There may be more than one complete set of commuting observables for a system. For example (see Section 5.5A), in the case of a two=particle system, the angular momenta of the individual particles and their components along a reference axis (together with the Hamiltonian) form one complete set, while the individual angular momenta, the total angular momentum and its component along the reference axis, form another complete set. The common eigenvectors of any of these sets, could be used for specifying the states of the system. Of course, the description in terms of the different sets are *equivalent*, since these different sets merely define different representations related by Unitary transformations (see, *Change of Basis*, Section 2.4).

Expectation Values and Probabilities

If we make a measurement of a physical observable, the outcome would be one of the eigenvalues of the corresponding operator. If the system is in a state represented by an eigenvector of the operator, then the eigenvalue obtained will be the eigenvalue belonging to this eigenvector. But suppose the system is not in a state corresponding to an eigenvector of the operator. What would be the result of the measurement then? The answer is provided by Postulate III.

Postulate III. *The outcome of the measurement of an observable of a quantum mechanical system in a particular state is given by the expectation value of the corresponding operator in that state.*

The expectation value $<\hat{A}>_{\mathbf{X}}$ of an operator $\hat{A}$ in the state **X**, is defined as,

$$<\hat{A}>_{\mathbf{X}} = \frac{(\mathbf{X}, \hat{A}\mathbf{X})}{(\mathbf{X}, \mathbf{X})} = \frac{<X \mid \hat{A} \mid X>}{<X \mid X>} \tag{3.2a}$$

If $|X>$ is an eigenvector of $\hat{A}$, say, $|X> \equiv |u_k>$, where

$$\hat{A} \mid u_k> = \alpha_k \mid u_k>,$$

then,

$$<\hat{A}>_X = \alpha_k, \tag{3.2b}$$

which is in agreement with our earlier statement. If $|X>$ is not an eigenvector of $\hat{A}$, then, $|X>$ can be expanded in terms of the eigenvectors $\{|u_k>\}$ of $\hat{A}$, which form a complete, orthonormal set (see property (*H*3) of Hermitian operators, Section 2.2):

$$|X> = \sum_k |u_k> <u_k \mid X>, \tag{3.3}$$

so that,

$$<\hat{A}>_X = \frac{\sum_k \alpha_k |<u_k \mid X>|^2}{\sum_k |<u_k \mid X>|^2} = \sum_k w_k \alpha_k \tag{3.4}$$

where

$$w_k = \frac{|< u_k \mid X >|^2}{\sum_k |< u_k \mid X >|^2} \tag{3.5}$$

Now, according to Postulate I, the outcome of any single measurement is an eigenvalue of the operator. Equation (3.4) should, therefore, be interpreted as giving the outcome of a large number of measurements under identical conditions, or, equivalently, the outcome of a measurement on a large number of identical (similarly-prepared) systems. That is, $< \hat{A} >_X$ is the *weighted average* of a number of measurements. Each measurement will yield one or the other eigenvalue of $\hat{A}$. $|< u_k \mid X >|^2$ is the *frequency* with which the eigenvalue α_k occurs in the measurement. The ratio of this frequency to the total number of measurements, $\sum_k |< u_k \mid X >|^2$, is the *weight* w_k (Eq.(3.5)) of the eigenvalue α_k in the measurement. w_k could be interpreted as the *probability* that a *single* measurement yields the value α_k. The result, $\sum_k w_k = 1$, reinforces this interpretation, since $\sum_k w_k$ is the probability that a single measurement yields one or the other eigenvalue of $\hat{A}$.

If the state vector $| X >$ is normalized, we have,

$$< X \mid X > \equiv \sum_k |< u_k \mid X >|^2 = 1, \tag{3.6}$$

so that,

$$w_k = |< u_k \mid X >|^2 . \tag{3.5a}$$

In this case, $| X >$ itself could be regarded as a probability amplitude. Unless otherwise specified, we will assume hereafter that $| X >$ is normalized.

Now, the probability that a measurement yields the value α_k, is the probability that the system is found in the state $| u_k >$. Therefore, $|< u_k \mid X >|^2$ is the probability that the system is in the state $| u_k >$. Hence, $< u_k \mid X >$ could be regarded as the *probability amplitude*[4] for the system to be found in $| u_k >$. Of course, this concept of the probability amplitude is based on the premise that $| X >$ as expressed by Eq. (3) has meaning as the state vector of a physical system. The justification for the premise comes from experiments on interference and diffraction phenomena (see Section 1.1). Taking the co-ordinate representation (see Section 2.A) of Eq (3.3), we get,

$$< \mathbf{r} \mid X > = \sum_k < u_k \mid X > \; < \mathbf{r} \mid u_k >,$$

4. This term is derived from wave optics, where the intensity of light at a point is given by the square of the amplitude of the wave at that point. The probability is defined as the absolute square because of the possibility that $< u_k \mid X >$ may be complex.

or,

$$\mathbf{X}(\mathbf{r}) = \sum_k x_k \mathbf{u}_k(\mathbf{r}), \tag{3.3a}$$

where

$$x_k = < u_k \mid X >. \tag{3.7}$$

Eq. (3.3a) is, obviously, equivalent to Eq. (1.3) and Eq. (1.11).

Does the physical system, of which $| X >$ is the state vector, represent a single particle? The answer is that it is possible. In that case, $| X >$ would be the eigenvector of some operator $\hat{H}$ that does not commute with $\hat{A}$, so that $| X >$ would be a state in which H has a definite value but A has no definite value. As an illustration, consider the linear harmonic oscillator. Let $| X >$ be an eigenvector of the operator corresponding to the Hamiltonian of the system. That is, $| X >$ is a state in which the oscillator has a definite energy. $\hat{A}$ could, then, correspond to the position co-ordinate of the oscillator. If x_0 is the amplitude of oscillation, any one measurement of A will give a value which would lie anywhere between $-x_0$ and $+x_0$. The average of a large number of repeated measurements would be zero, corresponding to the equilibrium position of the oscillator. The same average value for the position co-ordinate would, however, be obtained if the measurement is made on an ensemble of identical oscillators all of which are in the same energy state $| X >$. Thus, as far as measurement of $\hat{A}$ is concerned, $| X >$ could as well represent an ensemble of oscillators as a single oscillator. In the former case, $|< u_k \mid X >|^2$ is to be interpreted as proportional to the *number of oscillators* in the ensemble whose position co-ordinates are equal to that corresponding to $| u_k >$ at the time of measurement[5]. In the latter case, $|< u_k \mid X >|^2$ should be interpreted as proportional to the *number of measurements* which results in a value for the position co-ordinate that is appropriate to $| u_k >$. In either case, the measurement of A is of a statistical nature. The outcome of any single measurement cannot be predicted.

In the case of microscopic systems, the interpretation of $| X >$ as representing the state of an ensemble is to be preferred, since measurements are actually done on ensembles rather than on single systems. For example, it is not practical to isolate, say one hydrogen atom, and make repeated measurements on it.

In this context, it might be relevant to make a distinction between the state $| u_k >$ and the state $| X >$. We will refer to the former as a *pure state* and to the latter as

5. The position co-ordinates of more than one oscillator could be equal because these are measured from the equilibrium positions (which are different for different oscillators) of the oscillators.

a *mixed state*. When concerned with measurements, we will assume that a pure state corresponds to an individual system and a mixed state to an ensemble[6].

We, thus, see that Postulate III leads to a statistical interpretation of quantum mechanics. Alternatively, we could say that this postulate expresses the statistical nature of quantum mechanics. The predictions of quantum mechanics are also, therefore, of a statistical nature. For example, even if we are given complete data (in the quantum-mechanical sense) on a radioactive nucleus, we will not be able to predict, on the basis of quantum mechanics, *when* exactly it will disintegrate; what we can predict is the *number* of nuclei that will disintegrate in a given time, in a sample (or, ensemble) consisting of a large number of such nuclei. True to the statistical nature of the prediction, the predicted number would be nearer to the actual number, the larger the number of nuclei involved. The analogy of the number of deaths in a community of people might help to elucidate the point. It is impossible to predict when exactly a particular individual in the community will die, but we will be able to predict, on the basis of previous data, fairly accurately the number of deaths that will take place, say, during the coming month. The accuracy of the prediction could be increased by replacing the month by an year. Alternatively for the same period, the accuracy of the prediction would be more in the case of a city with a population of a few million people than in the case of a village with a few thousand people.

Lest this analogy lead to the mistaken notion that quantum mechanics is merely a statistical theory like, say, classical thermodynamics, let it be emphasized that the concept of the probability amplitudes that obey the principle of superposition (Eqs. (3.7) and (3.3a)), is a novel element in the theory. This results in a law of combination of probabilities for different but (experimentally) indistinguishable alternatives, that is quite foreign to classical statistics[7] (see Eq. (1.4)).

---

6. Note that the designation as pure and mixed states is with reference to a given observable, in this case $\hat{A}$. A mixed state of one observable could be a pure state of another observable. In our example of the oscillator, $|X>$ is a mixed state of $\hat{A}$, but a pure state of $\hat{H}$. This difference in the role of $|X>$ when referred to different observables, does not lead to any inconsistency, since we are concerned with the measurement of only A, whereas H is brought in only to *specify* $|X>$. If we were to talk of the measurement of both A and H, then $|X>$ would have to be the same type of state with respect to both $\hat{A}$ and $\hat{H}$ (see, Section 3.2).

7. In the classical theory, there are two ways in which probabilities for different events are combined. If P_A and P_B are, respectively, the probabilities for events A and B, then the probability P_{AB} for the combined event is given by either (i) $P_{AB} = P_A + P_B$, or (ii) $P_{AB} = P_A P_B$. In (i), P_{AB} is the probability for either A or B to happen. Thus if P_A is the probability for a coin in a toss to fall with head-up, and P_B to fall with tail-up, then $P_A + P_B$ is the probability to fall with either head or tail up. Thus P_A and P_B in this case represents probabilities for *mutually exclusive* events. In (ii) P_A and P_B represents probabilities for *independent* events. Thus P_A could be the probability for the coin to fall with head-up in one toss, and P_B the probability to fall with head-up in another toss. P_{AB} is, then, the probability that the coin falls with head-up in both the tossings. The important point is that, both in (i) and (ii), it is the probabilities that are to be added or multiplied and there is no such thing as a probability amplitude.

Quantum Mechanical Operators

We will now consider the question of obtaining the Hermitian operator that represents a given physical observable. We know that, in classical mechanics, a physical observable corresponds to a dynamical variable which could be constructed from pairs of basic, canonically conjugate variables such as generalized co-ordinates and momenta. For example, the dynamical variable corresponding to the total energy of a system is the Hamiltonian H which is a function of the generalized co-ordinates and the generalized momenta. The method to obtain the quantum mechanical operator from the classical dynamical variable, is the subject of Postulate IV:

Postulate IV: *The quantum mechanical operator corresponding to a dynamical variable is obtained by replacing the classical canonical variables in the latter by the corresponding quantum mechanical operators.*

For example, the Hamiltonian of a linear harmonic oscillator is given, in classical mechanics (in cartesian co-ordinates), by

$$H \equiv H(x,p) = \frac{p^2}{2m} + \frac{1}{2}Kx^2, \tag{3.8}$$

where m is the mass of the oscillator and K is a constant. Here, x (the position co-ordinate) and p (the linear momentum) are the basic canonic variables. According to the above postulate, the quantum mechanical operator corresponding to H is given by

$$\hat{H} \equiv \hat{H}(\hat{x},\hat{p}) = \frac{\hat{p}^2}{2m} + \frac{1}{2}K\hat{x}^2, \tag{3.9}$$

where $\hat{x}$ and $\hat{p}$ are Hermitian operators corresponding, respectively, to x and p.

While making the replacement of the canonical variables by quantum mechanical operators, care should be taken to preserve the proper order of the variables, since the operators need not commute. Care should also be taken to see that the resulting operator is Hermitian. For illustration, let us consider the (orbital) angular momentum; $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. We have,

$$L_x = yp_z - zp_y, \text{ cyclic,} \tag{3.10}$$

so that,

$$L_x^2 = yp_z yp_z + zp_y zp_y - yp_z zp_y - zp_y yp_z \tag{3.11a}$$

$$= y^2 p_z^2 + z^2 p_y^2 - 2y p_y z p_z \tag{3.11b}$$

Classically, both expressions (3.11a) and (3.11b) are correct. However, replacing the canonical variables in (3.11b) by the corresponding operators will lead to a wrong expression for $\hat{L}_x^2$. In fact, $\hat{L}_x^2$ so obtained would be non-Hermitian. This

difficulty could be avoided by making the simplifications only after the replacement of the variables by operators. Only those operators that commute should be permuted. Thus, the correct expression for $\hat{L}_x^2$ is obtained by replacing the variables by corresponding operators in Eq. (3.11a).

As another example, consider a dynamical variable C which is the product of two other dynamical variables A and B. That is, $C = AB$. Then, the Hermitian operator corresponding to C is not $\hat{A}\hat{B}$, since if $\hat{A}\hat{B} \neq \hat{B}\hat{A}$, then $\hat{A}\hat{B}$ would not be Hermitian. We should construct a combination of $\hat{A}$ and $\hat{B}$ that is Hermitian. Such a combination is given by $\hat{C} = \frac{1}{2}(\hat{A}\hat{B} + \hat{B}\hat{A})$.

Postulate IV is of no use in the case of observables like spin, isospin, etc., which have no classical analogue. In such cases, other considerations, such as the algebra they obey, might help to define the operator (see Section 5.1).

The question now remaining to be answered is how to specify the quantum mechanical operators corresponding to the basic canonical variables. Answer to this question is provided by our final postulate.

Postulate V: *Any pair of canonically conjugate operators will satisfy the following Heisenberg commutation rules :*

$$[\hat{q}_i, \hat{q}_k] = \hat{0} = [\hat{p}_i, \hat{p}_k] \tag{3.12a}$$

$$[\hat{q}_i, \hat{p}_k] = i\hbar\hat{1}\delta_{ik} = i\hbar\hat{\delta}_{ik} \tag{3.12b}$$

Here, $\hat{q}_i$ is the operator corresponding to the generalized co-ordinate q_i while $\hat{p}_i$ is the operator corresponding to the generalized momentum p_i that is canonically conjugate[8] to q_i.

Thus, if the q_i are cartesian co-ordinates, then the p_i are the components of linear momentum. If q_i are angles, p_i are components of angular momentum, and so on.

---

8. The fact that q_i and p_i are canonically conjugate to each other is expressed in classical mechanics, in any of the following equivalent ways:
   (i) They satisfy Hamiltonian's canonical equations,
   $$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}; \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i},$$
   where $H(q, p, t)$ is the Hamiltonian of the mechanical system.
   (ii) They satisfy the Poisson brackets,
   $$\{q_i, p_k\} = \delta_{ik};$$
   $$\{q_i, q_k\} = 0 = \{p_i, p_k\}.$$
   (iii) They satisfy the relationship,
   $p_i = \dfrac{\partial S}{\partial q_i}$, where S is the action (see Eq. (1.14b)).

Postulate V may be called the *axiom of quantization.* For, any operator that has a classical analogue, can be expressed as a function of the q_i and the p_i, and the quantal properties of such an operator would follow from the relations (3.12a, b). For example, the substitution of the operators $\hat{y}, \hat{p}_y, \hat{z}, \hat{p}_z$, etc. in place of the variables $y, p_y, z, p_z \ldots$ in Eq. (3.10) leads to the following relations for the component of $\hat{L}$:

$$[\hat{L}_x, \hat{L}_y] = i\hbar \hat{L}_z, \text{ cyclic,} \tag{3.13}$$

where, use has been made of Eqs. (3.12a, b). It is shown, in the theory of angular momentum (see Section 5.2), that Eq. (3.13) completely determines the eigenvalues and eigenvectors and, hence, the quantal properties of angular momentum.

Of course, the designation of Postulate V as an axiom of quantization, is meaningful only if the commutation relations (3.12) are invariant under Unitary transformations (that is, under change of basis–see Eq. (2.124)). This is, indeed, the case, as is easily proved.

The following commutation relations, which are obtained from the basic commutation relations (3.12b) by induction, might be of help in many applications:

$$[\hat{q}, \hat{p}^n] = i\hbar \frac{\partial}{\partial \hat{p}}(\hat{p}^n), \tag{3.14a}$$

$$[\hat{p}, \hat{q}^n] = -i\hbar \frac{\partial}{\partial \hat{q}}(\hat{q}^n) \tag{3.14b}$$

$$[\hat{q}, \hat{A}(\hat{q}, \hat{p})] = i\hbar \frac{\partial \hat{A}}{\partial \hat{p}}, \tag{3.14c}$$

$$[\hat{p}, \hat{A}(\hat{q}, \hat{p})] = -i\hbar \frac{\partial \hat{A}}{\partial \hat{q}}, \tag{3.14d}$$

where, $\hat{A}$ is a function of $\hat{q}$ and $\hat{p}$.

These relations suggest the identities,

$$\hat{p} \equiv -i\hbar \frac{\partial}{\partial \hat{q}}, \tag{3.15a}$$

and

$$\hat{q} \equiv i\hbar \frac{\partial}{\partial \hat{p}}. \tag{3.15b}$$

Problem 3.1: Prove the relations (3.14a-d).

Explicit Representation of Operators

The question of how to represent the operators *explicitly*, is still left. So far we have defined the canonical operators only by the algebra which they should satisfy. The explicit form of the operators will depend upon the type of Hilbert space chosen and the representation selected. In the Schrödinger method the Hilbert space is a continuous one, that is, a function-space. We can still

choose either the co-ordinate or the momentum representation. In the former, the state vectors are functions of the position co-ordinates and the generalized co-ordinates are, therefore, multiplicative operators :

$$\hat{q}_i \equiv q_i \tag{3.16^1}$$

The relation (3.12b), $[\hat{q}_i, \hat{p}_k] = i\hbar\hat{\delta}_{ik}$, then requires that the conjugate momenta be differential operators:

$$\hat{p}_i \equiv -i\hbar\,(\mathrm{grad}_q)_i, \tag{3.17^1}$$

where, the gradient is with respect to the position co-ordinates. Thus, in the cartesian system, the canonical operators are,

$$\hat{x} = x;\ \hat{p}_x = -i\hbar\frac{\partial}{\partial x};$$

$$\hat{y} = y;\ \hat{p}_y = -i\hbar\frac{\partial}{\partial y};$$

$$\hat{z} = z;\ \hat{p}_z = -i\hbar\frac{\partial}{\partial z}. \tag{3.18}$$

This gives, for the Hamiltonian of linear harmonic oscillator (Eq. (3.9)), the explicit expression,

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}Kx^2, \tag{3.9^1}$$

and for the components of the orbital angular momentum (Eq. (3.10)),

$$\hat{L}_x = -i\hbar\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right), \text{ cyclic} \tag{3.10^1}$$

If we choose the momentum representation, then the state vectors are functions of the momenta. The momenta are then represented by multiplicative operators:

$$\hat{p}_i \equiv pi, \tag{3.17^2}$$

and the co-ordinates by differential operators.

$$\hat{q}_i \equiv i\hbar\,(\mathrm{grad}_p)_i, \tag{3.16^2}$$

where the gradient is with respect to the momenta.

Again, in the cartesian system, the canonical operators are:

$$\hat{p}_x = p_x;\ \hat{x} = i\hbar\frac{\partial}{\partial p_x};$$

$$\hat{p}_y = p_y;\ \hat{y} = i\hbar\frac{\partial}{\partial p_y};$$

$$\hat{p}_z = p_z;\ \hat{z} = i\hbar\frac{\partial}{\partial p_z}. \tag{3.18^1}$$

Hence, corresponding to Eqs. (3.9) and (3.10), we have,

$$\hat{H} = \frac{p_x^2}{2m} - \frac{1}{2}\hbar^2 K \frac{d^2}{dp_x^2}, \tag{3.9^2}$$

and

$$\hat{L}_x = i\hbar\left(\frac{\partial}{\partial p_y} p_z - \frac{\partial}{\partial p_z} p_y\right), \text{ cyclic} \tag{3.10^2}$$

Problem 3.2: Show that Eq. (3.10^2) is consistent with the communication relations (3.13).

In the Heisenberg method, the Hilbert space is discrete (but infinite-dimensional). The operators are represented by (infinite-dimensional) discrete matrices. For example, in the case of the linear harmonic oscillator,

$$\hat{x} = \frac{\alpha}{\sqrt{2}}\begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & . & . \\ \sqrt{1} & 0 & \sqrt{2} & 0 & . & . \\ 0 & \sqrt{2} & 0 & \sqrt{3} & . & . \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} & 0 \\ 0 & 0 & 0 & \sqrt{4} & 0 & \sqrt{5} \end{pmatrix}, \tag{3.16^3}$$

and

$$\hat{p}_x = -i\hbar\frac{1}{\alpha\sqrt{2}}\begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & . & . & . \\ -\sqrt{1} & 0 & \sqrt{0} & . & . & . & \\ 0 & -\sqrt{2} & 0 & \sqrt{3} & 0 & . & . \\ 0 & 0 & -\sqrt{3} & 0 & \sqrt{4} & 0 & . \\ 0 & 0 & 0 & -\sqrt{4} & 0 & \sqrt{5} & . \end{pmatrix} \tag{3.17^3}$$

That is,

$$x_{mn} = \left(\frac{\alpha}{\sqrt{2}}\right)\sqrt{n}, \text{ if } m = n+1,$$

$$= \frac{\alpha}{\sqrt{2}}\sqrt{n-1}, \text{ if } m = n-1, \tag{3.19}$$

$$= 0, \text{ otherwise,}$$

and,

$$(p_x)_{mn} = -i\hbar\left(\frac{1}{\alpha\sqrt{2}}\right)\sqrt{n}, \text{ if } m = n+1$$

$$= i\hbar\left(\frac{1}{\alpha\sqrt{2}}\right)\sqrt{n-1}, \text{ if } m = n-1 \tag{3.20}$$

$$= 0, \text{ otherwise,}$$

where

$$\alpha^2 = \frac{\hbar}{\sqrt{mK}} = \frac{\hbar}{m\omega}, \tag{3.21}$$

$\omega = \sqrt{\frac{K}{m}}$, being the frequency of the oscillator.

Substituting Eqs. (3.19) and (3.20) in Eq. (3.9), we get,

$$\hat{H} = \hbar\omega \begin{pmatrix} 1/2 & 0 & 0 & 0 & . & . & . \\ 0 & 3/2 & 0 & 0 & . & . & . \\ 0 & 0 & 5/2 & 0 & 0 & . & . \\ 0 & 0 & 0 & 7/2 & 0 & . & . \\ 0 & 0 & 0 & 0 & 9/2 & 0 & . \\ . & . & . & . & . & . & . \end{pmatrix}, \tag{3.9^3}$$

or
$$H_{mn} = E_n \delta_{mn}, \tag{3.22a}$$

where
$$E_n = \left(n + \frac{1}{2}\right)\hbar w, \; n = 0, 1, 2, \ldots \tag{3.22b}$$

Problem 3.3: Show that expressions (3.16^3) and (3.17^3) are consistent with the commutation relation, $[\hat{x}, \hat{p}_x] = i\hbar\hat{1}$.

Problem 3.4: Show that the linear vector space, on which the operators $\hat{q}_i$ and $\hat{p}_i$ in Eq. (3.12b) are defined, is infinite-dimensional.

3.2 THE UNCERTAINTY PRINCIPLE

We shall now discuss the result of trying to measure experimentally the values of two incompatible (or, complementary) variables for a physical system which is in a quantum state represented by the vector ψ. Such variables, as we have seen (see Postulate II), correspond to noncommuting Hermitian operators, say $\hat{A}$ and $\hat{B}$. That is,

$$[\hat{A}, \hat{B}] = i\hat{C}, \tag{3.23}$$

where $\hat{C}$ is Hermitian and non-null.

We will assume that $|\psi>$ is normalized ; that is,

$$<\psi|\psi> = 1. \tag{3.24}$$

First we will discuss the type of state vector for which it is meaningful to talk of the values of incompatible variables.

Case 1: ψ is an eigenvector of one of the operators, say $\hat{A}$.

According to Postulates I and III, then, a measurement of A will yield the value α which is the eigenvalue of $\hat{A}$ belonging to the eigenvector ψ. The attempt to measure B will carry the system over to an eigenvector of $\hat{B}$, which is not an eigenvector of $\hat{A}$ since $\hat{A}\hat{B} \neq \hat{B}\hat{A}$. Therefore, the measured value of B would refer to a state which is different from the one represented by ψ. In other words, it is not possible to specify values for both of two incompatible variables when the system is in a state corresponding to an eigenvector of one of the two operators that represent the variables[9].

---

9. It might be argued, on the basis of Postulate III, that the measurement of B should result in the expectation value $<\hat{B}>_\psi$. But this would require that we interpret ψ as a mixed state (see footnote 6) representing an ensemble. On the other hand, in the case of A, we have to interpret ψ as a pure state representing an individual system or, alternatively, an ensemble in which all the individual members are in the same quantum state. The inconsistency is avoided by assuming that only one of the two variables A and B can be measured in the state ψ.

Case 2: ψ is not an eigenvector of either $\hat{A}$ or $\hat{B}$. That is ψ is a mixed state.

According to Postulate III, then, the measurement of A gives the expectation value $<\hat{A}>_\psi$ while the measurement of B yields the number $<\hat{B}>_\psi$. Thus, it is possible to specify values for both A and B in the state ψ without having to interpret $|\psi>$ differently in the two cases. In the remainder of this Section, we will assume that ψ is of this type. We will also drop the subscript ψ from the expectation values, as we are dealing with only one state-vector.

The General Uncertainty Relationship

Now, an expectation value is an average over several measurements. The individual measurements will deviate from the average value, some on the lower side and some others on the upper side. The average of these deviations would, of course, be zero. But the average of the squares of the deviations (called 'mean-square-deviation') would be non-zero. The square root of the mean-square-deviation, referred to as the *root-mean-square* (or *standard*) deviation, could be taken as a measure of the 'spread' in the measured values. These spreads ΔA and ΔB in the measured values of A and B are called the *uncertainties* in the measurement of A and B. Thus, the uncertainties ΔA and ΔB associated with the measurements of A and B in a given state are given by[10]

$$\Delta A = \{<(\hat{A} - <\hat{A}>)^2>\}^{1/2}$$

$$= \{<\psi|(\hat{A} - <\hat{A}>)^2|\psi>\}^{1/2}$$

$$= \{<\psi(\hat{A} - <\hat{A}>)|(\hat{A} - <\hat{A}>)\psi>\}^{1/2}, \text{ since } \hat{A} \text{ is Hermitian;}$$

$$= ||\psi_1||, \tag{3.25a}$$

where $\psi_1 = (\hat{A} - <\hat{A}>)\psi$; (3.26a)

Similarly, $\Delta B = ||\psi_2||$, (3.25b)

with $\psi_2 = (\hat{B} - <\hat{B}>)\psi$ (3.26b)

The fact that ψ is a vector in the Hilbert space, results in a certain correlation between ΔA and ΔB. It is this correlation that is referred to variously as the *uncertainty relationship*, the *uncertainty principle* or the *principle of indeterminacy*[11]. Whereas there is divergence of opinion among physicists as to the meaning of the relationship (see, footnote 27, Chapter 1), there is agreement on the relationship itself. A derivation of the relationship follows:

According to Schwarz inequality (Eq. (2.14)), we have, since $|\psi_1>$ and $|\psi_2>$ are vectors in the Hilbert space,

---

10. Note that ΔA would have been zero if ψ were an eigenvector of $\hat{A}$.
11. Ballentine, L. E. (Ref. 9, Chapter 1) suggests the name *statistical dispersion principle*.

$$||\psi_1|| \cdot ||\psi_2|| \geq |<\psi_1 | \psi_2>| \tag{3.27a}$$

$$\geq |\ \mathrm{Im}\ <\psi_1 | \psi_2>| \tag{3.27b}$$

$$\geq |\frac{1}{2i}(<\psi_1 | \psi_2> - <\psi_2 | \psi_1>)|,$$

since $$<\psi_1 | \psi_2>^* = <\psi_2 | \psi_1>.$$

That is

$$\Delta A \cdot \Delta B \geq \frac{1}{2}\left|\frac{1}{i}\{<\psi(\hat{A} - <\hat{A}>) | (\hat{B} - <\hat{B}>)\psi> - <\psi | (\hat{B} - <\hat{B}>) | (\hat{A} - <\hat{A}>)\psi>\}\right|$$

$$\geq \frac{1}{2}|\frac{1}{i}<\psi | (\hat{A}\hat{B} - \hat{B}\hat{A}) | \psi>|$$

$$\geq \frac{1}{2}|\frac{1}{i}<[\hat{A}, \hat{B}]>|$$

$$\geq \frac{1}{2}|<\hat{C}>|. \tag{3.28}$$

Putting $\hat{A} = \hat{q}_j$ and $\hat{B} = \hat{p}_k$ in Eq. (3.23), we have, from Eq. (3.12b),

$$\hat{C} = \hbar\hat{\delta}_{jk},$$

so that Eq. (3.28) reads,

$$\Delta q_j \cdot \Delta p_k \geq \left(\frac{\hbar}{2}\right)\delta_{jk} \tag{3.29}$$

Thus, if q_j are the components of the position vector **r** and p_k the components of the linear momentum **p** of a particle, then, we have

$$\Delta x \cdot \Delta p_x \geq \hbar/2, \text{ cyclic}. \tag{3.29a}$$

Inequality (3.29a) is known as *Heisenberg's Uncertainty relationship*. (3.29) is a generalization of (3.29a) to the case of all pairs of canonically conjugate variables while (3.28) is a generalisation to any pair of dynamical variables that correspond to non-commuting operators[12]. Thus, if $p_k = L_z$, the z-component of angular momentum, then $q_k = \phi$, the azimuthal angle, so that inequality (3.29) requires that

$$\Delta\phi \cdot \Delta L_z \geq \frac{\hbar}{2}. \tag{3.29b}$$

Similarly, putting, $\hat{A} = \hat{L}_x$ and $\hat{B} = \hat{L}_y$, in Eq. (3.23), we have from Eq. (3.13), $\hat{C} = \hbar\hat{L}_z$. Hence, from (3.28), we get,

---

12. The generalized uncertainty relationship (3.28) was first derived by H.P. Robertson [*Phys. Rev.*, **34**, 163 (1929)]

$$\Delta L_x \cdot \Delta L_y \geq \frac{\hbar}{2} |< \hat{L}_z >| \tag{3.28a}$$

Note that the inequality (3.28) is independent of $< \hat{A} >$ and $< \hat{B} >$, so that the uncertainty relationship holds even for states in which the expectation values of $\hat{A}$ and $\hat{B}$ are zero. An energy eigenstate of a linear harmonic oscillator (Hamiltonian given by Eq. (3.9)) is an example :

Let ψ represent an eigenvector of $\hat{H}$ belonging to the eigenvalue E. Also, let $\hat{A}$ and $\hat{B}$ be, respectively, the operators corresponding to position and momentum. That is,

$$\hat{A} \equiv \hat{x}; \hat{B} \equiv \hat{p}.$$

From, Eqs. (3.25a) and (3.25b), we have,

$$\Delta x = \{< \hat{x}^2 >\}^{1/2}; \Delta p = \{< \hat{p}^2 >\}^{1/2} \tag{3.30}$$

Since both x^2 and p^2 are positive for the oscillator, we have,

$$\Delta x > 0; \Delta p > 0,$$

so that,

$$\Delta x \cdot \Delta p > 0. \tag{3.31}$$

In fact, according to (3.29a), we should have,

$$\Delta x \cdot \Delta p \geq \hbar/2. \tag{3.29a'}$$

An interesting consequence of (3.31) is that the lowest energy, referred to as the 'zero-point energy' of the oscillator is different from zero. An estimate of this zero-point energy could be made using the expressions (3.9), (3.30) and (3.29a '). We have,

$$E = < \psi | \hat{H} | \psi > \equiv < \hat{H} > = \frac{1}{2m} < \hat{p}^2 > + \frac{1}{2} K < \hat{x}^2 >$$

$$= \frac{1}{2m} (\Delta p)^2 + \frac{K}{2} (\Delta x)^2$$

$$\geq \frac{1}{2m} (\Delta p)^2 + \frac{\hbar^2 K}{8.} \frac{1}{(\Delta p)^2}. \tag{3.32}$$

The minimum value of E is given by the minimum of the expression,

$$\frac{1}{2m} (\Delta p)^2 + \frac{\hbar^2 K}{8} \cdot \frac{1}{(\Delta p)^2},$$

That is,

$$E_{min} = E_0 = \left[\frac{(\Delta p)^2}{2m} + \frac{\hbar^2 K}{8(\Delta p_0)^2} \right]_{min}$$

$$= \frac{(\Delta p_0)^2}{2m} + \frac{\hbar^2 K}{8(\Delta p_0)^2}, \tag{3.33a}$$

where,

$$(\Delta p_0)^2 = \frac{1}{2}\sqrt{\hbar^2 K m} = \left(\frac{m}{2}\right)\hbar\omega_c, \tag{3.34}$$

ω_c being the classical frequency of the oscillator, given by,

$$\omega_c = \left(\frac{K}{m}\right)^{1/2} \tag{3.35}$$

Thus,

$$E_0 = \frac{1}{2}\hbar\omega_c \tag{3.33b}$$

Another example is provided by a state-vector ψ which is a normalized eigenvector of $\hat{L}_z$, where $\hat{A}$ and $\hat{B}$ are to be identified, respectively, with $\hat{L}_x$ and $\hat{L}_y$. Since $\hat{L}_x$ and $\hat{L}_y$ do not commute with $\hat{L}_z$ (see Eq. (3.13)), ψ is an eigenvector of neither $\hat{L}_x$ nor $\hat{L}_y$, so that

$$<\hat{L}_x>_\psi = 0 = <\hat{L}_y>_\psi .$$

Hence, from Eq. (3.25a), we have,

$$(\Delta L_x)^2 = <\hat{L}_x^2>_\psi ; (\Delta L_y)^2 = <\hat{L}_y^2>_\psi \tag{3.36}$$

If[13] $<\hat{L}_z>_\psi = m\hbar$, we get, from (3.28a),

$$\Delta L_x \cdot \Delta L_y \geq \frac{1}{2} m\hbar^2. \tag{3.28b}$$

Relationships (3.36) and (3.28b) could be used to deduce the value of $<\hat{L}^2>_\psi$, as is done in the following problem, where,

$$\hat{\mathbf{L}}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2. \tag{3.37}$$

Problem 3.5: If ψ is the normalized eigenvector of $\hat{L}_z$ corresponding to the maximum value of $<\hat{L}_z>$, and if this maximum value is equal to $l\hbar$, show, using (3.28a), that $<\hat{\mathbf{L}}^2>_\psi = l(l+1)\hbar^2$.

The Minimum Uncertainty Product

Relationship (3.28), with the equal sign, is referred to as the minimum uncertainty product. Whether the equality sign in (3.28) is applicable or not depends on the nature of the state vector ψ. In other words, only if the state vector satisfies certain conditions, the minimum uncertainty product is realizable. We will now discuss these conditions on the state vector ψ.

Now, the equality in (3.28) requires equality in both (3.27a) and (3.27b). In the case of (3.27a), this requires that ψ_1 and ψ_2 be proportional to each other (see (2.14)), whereas in the case of (3.27b), the condition is that the real part of $<\psi_1 | \psi_2>$ be zero. Thus, we should have,

13. Angular momentum is measured in units of $\hbar$, in quantum mechanics.

$$\psi_1 = c\psi_2, \tag{3.38a}$$

and

$$<\psi_1 \mid \psi_2> + <\psi_2 \mid \psi_1> \ = 0. \tag{3.38b}$$

In order to illustrate how conditions (3.38a, b) restrict the form of ψ, let us take the case of the linear harmonic oscillator.

Taking $\hat{A} \equiv \hat{x}$; $\hat{B} \equiv \hat{p}$, Eq. (3.38a) reduces to (see Eqs. (3.26a, b)),

$$(\hat{x} - <\hat{x}>)\psi = c(\hat{p} - <\hat{p}>)\psi,$$

or,

$$\hat{x}\psi = c\,\hat{p}\psi, \tag{3.39}$$

since $<\hat{x}> \ = 0 = \ <\hat{p}>.$

Eq. (3.38b) becomes,

$$<\psi \mid (\hat{x}\hat{p} + \hat{p}\hat{x})\psi> \ = 0.$$

That is (using $\hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar\hat{1}$),

$$<\psi \mid \hat{x}\hat{p} \mid \psi> \ = (i\hbar/2). \tag{3.40}$$

From Eqs. (3.39) and (3.40), we get,

$$<\psi \mid \hat{x}^2 \mid \psi> \equiv (\Delta x)^2 = \frac{i\hbar c}{2}. \tag{3.41a}$$

But $(\Delta x)^2$ is real and positive, so that c is negative imaginary :

$$c = -ia, \text{ where, } a > 0; \tag{3.42}$$

and,

$$(\Delta x)^2 = (a\hbar/2). \tag{3.41b}$$

Substituting (3.42) in (3.39), we have,

$$\hat{x}\psi = -ia\,\hat{p}\psi. \tag{3.39a}$$

Taking the co-ordinate representation of this equation, we get (Eq. (3.18)),

$$x\psi(x) = -ia\left(-i\hbar\frac{d}{dx}\right)\psi(x),$$

or

$$\frac{d\psi}{\psi} = -(x/a\hbar)dx,$$

so that

$$\psi(x) = N \exp\,[-x^2/2a\hbar] \tag{3.43a}$$

$$= N \exp\left[\frac{-x^2}{4(\Delta x)^2}\right],$$

using (3.41b), where, N is normalizing constant to be determined from the relation,

$$< \psi \mid \psi > \equiv \int_{-\infty}^{+\infty} | \psi(x) |^2 dx = 1. \tag{3.44}$$

Eqs. (3.41a) and (3.42) yield, since

$$\int_{-\infty}^{+\infty} e^{-b^2x^2} dx = \frac{\sqrt{\pi}}{b},$$

$$N = \left[\frac{1}{2\pi(\Delta x)^2}\right]^{1/4}, \tag{3.45}$$

$$\psi(x) = [2\pi(\Delta x)^2]^{-1/4} \exp\left[-\frac{x^2}{4(\Delta x)^2}\right]. \tag{3.43b}$$

If we had used the momentum representation[14], we would have got,

$$\phi(p) = [2\pi(\Delta p)^2]^{-1/4} \exp\left[-\frac{p^2}{4(\Delta p)^2}\right], \tag{3.43c}$$

with $$(\Delta p)^2 = (\hbar/2a). \tag{3.41c}$$

Thus, ψ is *Gaussian*. A plot of $\psi(x)$ against x is given in Fig. 3.1.

The Time-Energy Uncertainty Relationship

Does there exist a time-energy uncertainty relationship analogous to the position-momentum uncertainty relationship (3.29a)? In other words, is it possible to substitute t for q_j and E for p_j in (3.29)? An affirmative answer to the question would imply the following premises:

(i) Time and energy (t and H) are complementary (or, conjugate) variables.

(ii) A Hermitian operator corresponding to t can be defined.

Unfortunately, justifying either of the premises proves to be difficult. The reason is that time has a dual role in mechanics: as a parameter and as a dynamical variable. And it is in its role as a parameter that it makes its appearance most of the time, especially in classical mechanics. Even when t plays the role of a dynamical variable, as in Eq. (1.14a), it appears to be conjugate, if at all, to $-H$ rather than to H (see criterion (iii) listed in footnote 8 of this chapter). Assuming, then, that t and $-H$ are conjugate variables, the operator $\hat{t}$ corresponding to time in an energy representation should be given by (cf. Eqs. (3.15b) and (3.18[1])),

$$\hat{t} = -i\hbar \frac{\partial}{\partial H} = -i\hbar \frac{\partial}{\partial E}. \tag{3.46}$$

---

14. Note that the co-ordinate and momentum representations of $| \psi >$ are given, respectively, by

$$\psi(x) = < x \mid \psi > \text{ and}$$

$$\phi(p) = < p \mid \psi > = \left(\frac{1}{2\pi\hbar}\right)^{1/2} \int \psi(x) \exp\left(-\frac{i}{\hbar} px\right) dx$$

[see Eqs. (2.128a), (2.129a) and (2.140a)]

However, it turns out that $\hat{t}$ given by Eq. (3.46) is not Hermitian.[15] It appears, therefore, that a time-energy uncertainty relationship, at least in non-relativistic quantum theory, has no place, within the frame work of (3.29)[16]. Nevertheless, a relationship of the form,

$$\Delta t \cdot \Delta E \geq \hbar, \tag{3.47}$$

is cited in many books on quantum mechanics as a time-energy uncertainty relationship. The tenuous status of the semiequality (3.47) as a time-energy uncertainty relationship is, however, reflected not only in the fact that it is interpreted differently from (3.29), but also in the fact that (3.47) itself is given more than one interpretation.[17]

Relationship (3.47) could, however, be ligitimised as a time-energy uncertainty relationship by identifying the position-momentum uncertainty relationship, not with (3.29a), but with a relationship analogous to (3.47) based on equally vague premises such as that of a wave packet (see, J.L. Powell and B. Crasemann, footnote 15)[18]. But, in this book we will identify the uncertainty relationship with (3.28), so that (3.47) would not be regarded as an uncertainty relationship.[19]

It should be added here that, whereas (3.47) is of dubious validity as a time-energy uncertainty relationship, the relationship (3.47) itself has a legitimate place in quantum mechanics. Thus, in the case of radioactive decay, we have the relation,

$$\tau \Gamma \approx \hbar, \tag{3.47a}$$

where, τ is the *mean life* and Γ the *width* of the level. Similarly, in the case of transitions (between levels) induced by an external, constant perturbing field, we have (see Eq. (8. 169b))

$$T \cdot \Delta E \approx \hbar \tag{3.47b}$$

where, T is the duration of the field and ΔE is the separation between the levels.

---

15. See, Allock, G.R., *Ann. Phys.*, **53**, 253 (1969).
16. The same conclusion is arrived at, by a different procedure, in Ref. 10 of Chapter 1.
17. See, for example, Landau, L.D. and Lifshitz, E.M., *Quantum Mechanics* (Pergamon Press, Oxford 1965), II Edition, Section 44 ; Powell, J.L. and Crasemann, B. *Quantum Mechanics* (B.I. Publications, Delhi 1971), II Edition, Section 3.4; Messiah, A., *Quantum Mechanics* (North-Holland Publishing Co., Amsterdam 1961), Vol. 1, Chapter IV.
18. In this connection, see footnote 27 of Chapter 1.
19. Since, however, the last word on this subject has not been said (judging from the profusion of papers still appearing on the topic), the reader is referred to the current literature in addition to the following papers : Rayski, J, and Rayski, J.M., Jr., **On the meaning of the time-energy uncertainty relation;** Recamé, E., **A time operator and the time-energy uncertainty relation** [both in *The Uncertainty Principle and the Foundations of Quantum Mechanics* (John Wiley, London 1977)]; Bauer, M. and Mellow, P.A., *Ann. Phys.*, **111**, 38 (1978); Sörkin, R. *Found Phys.*, **9**, 123 (1979); de la Pena, Luis, *Amer. J. Phys.* **48**, 775 (1980); Srinivas, M.D. and Vijayalakshmi, R. *Pramana*, **16**, 173 (1981).

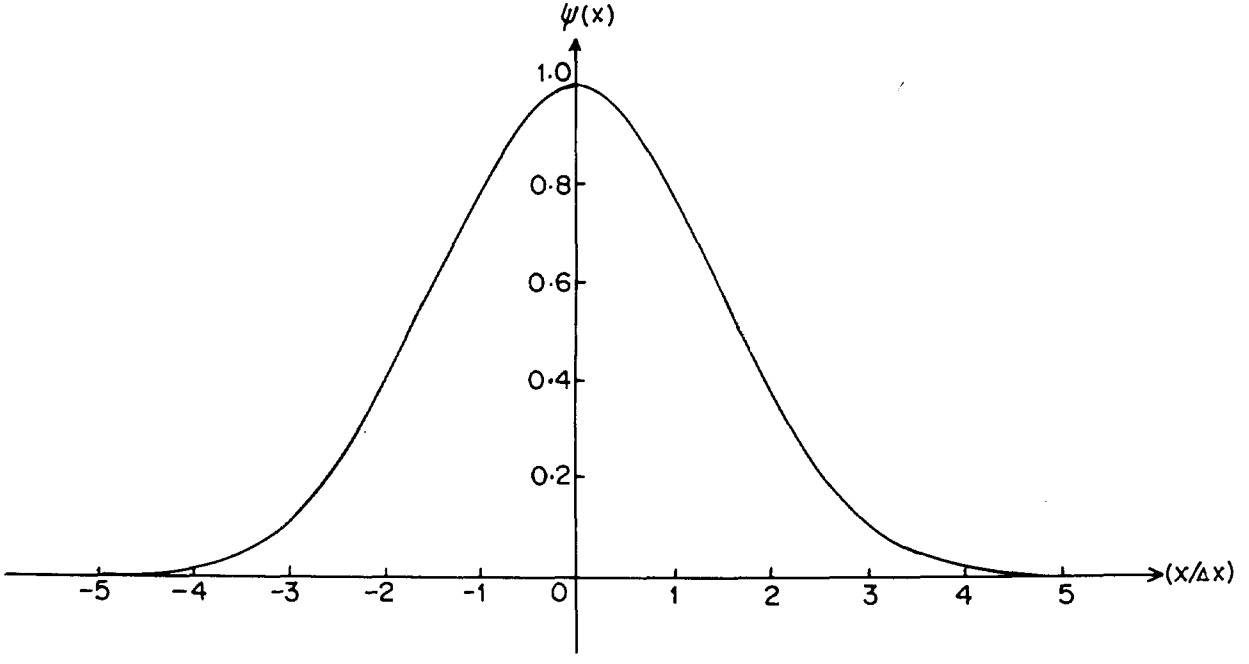


Fig. 3.1. Harmonic oscillator state-vector corresponding to the minimum uncertainty product.

3.3 DENSITY MATRIX

In connection with Postulate III, we have referred to a state represented by a basis vector $|u_k>$ as a *pure state* and a state represented by a *coherent mixture* $|X>$ of such states as a *mixed state*. Conventionally, however, a state represented by $|X>$ itself is called a pure state. A mixed state is then an *incoherent mixture* of such pure states. In other words, any state described by a state vector (wave function) as per Postulate II is a pure state while a state which cannot be described so is a mixed state. The question then arises how one will represent or characterise a mixed state. This is where the concept of *density matrix* or *statistical operator* comes in[20]. As the name implies, the density matrix is an operator, and not a vector.

Mixed States

Let us denote the pure states (assumed to be normalized) by $|X^{(i)}>$, where, in terms of the basis states, we have,

$$|X^{(i)}> = \sum_{k=1}^{n} c_k^{(i)} |u_k> . \tag{3.48}$$

The expectation value of an operator $\hat{A}$ in the state $|X^{(i)}>$ is then given by

$$<\hat{A}>_i = \sum_{j,k} c_j^{(i)*} c_k^{(i)} A_{jk}, \tag{3.49}$$

with A_{jk} given by Eq. (2.117a). By definition a mixed state is an incoherent mixture of the $|X^{(i)}>, (i = 1, 2, \ldots N)$ with *statistical weights* w_i such that $\sum_{i=1}^{N} w_i = 1$.

This means that the average value of A in the mixed state, denoted by $<\overline{\hat{A}}>$, is given by the expression

$$<\overline{\hat{A}}> = \sum_{i=1}^{N} w_i <\hat{A}>_i . \tag{3.50^1}$$

Substituting for $<\hat{A}>_i$ from Eq. (3.49), Eq. (3.50^1) reduces to

$$<\overline{\hat{A}}> = \mathrm{Tr}\,(\rho A), \tag{3.50^2}$$

where

$$\rho_{kj} = <u_k|\hat{\rho}|u_j> = \sum_i w_i c_k^{(i)} c_j^{(i)*} = \sum_i w_i <u_k|X^{(i)}><X^{(i)}|u_j>, \tag{3.51}$$

or

$$\hat{\rho} = \sum_i w_i |X^{(i)}><X^{(i)}| . \tag{3.52}$$

---

20. von Neumann, J. [*Gottinger Nachr*, 1, 246 (1927)] was the first to introduce this in physics.

The matrix ρ defined by Eq. (3.51) is the density matrix. The corresponding operator, Eq. (3.52), is called the density or statistical operator. As seen from the definition, the density matrix depends only on the pure states involved in the mixed state and their statistical weights in the letter. ρ, thus, truly characterises the mixed state.

Properties of the Density Matrix

The properties of the density matrix follow from its defining equations (3.50^2), (3.51) and (3.52). Thus, putting $\hat{A} = \hat{1}$ in (3.50^2), we have,

$$\mathrm{Tr}\,(\rho) = \overline{<\hat{1}>} = 1. \tag{3.53}$$

This follows from Eq. (3.51) as well, since $\sum_{k=1}^{n} |c_k^{(i)}|^2 = 1$ and $\sum_{i=1}^{N} w_i = 1$. This equation also shows that

$$\rho_{kj} = \rho_{jk}^{*}, \tag{3.54}$$

or that $\hat{\rho}$ is *Hermitian*.

Pure States:

A pure state could be looked upon as a mixed state characterised by the statistical weight $w_j = \delta_{ij}$. Then from (3.52) we see that

$$\hat{\rho}^2 = \hat{\rho} \tag{3.55}$$

The matrix elements of ρ in this case are given, according to (3.51), by

$$\rho_{kj} = c_k^{(i)} c_j^{(i)*}. \tag{3.56}$$

Problem 3.6: Show that Eq. (3.55) is a necessary as well as a sufficient condition for a pure state.

Eq. (3.55) shows that $\hat{\rho}$ is a projection operator with eigenvalues 1 and 0. The expectation value of an operator $\hat{A}$, in this case also, is given [according to Eqs. (3.49) and (3.56)] by

$$<\hat{A}> \; = \mathrm{Tr}\,(\rho A) \tag{3.50^3}$$

which is similar to Eq. (3.50^2). Thus, a pure state can be represented by a density matrix as well as by a state vector. The density matrix has the additional capability of representing a mixed state which cannot be represented by a state vector.

Two-level System

As an illustrative example, let us consider a 2-level system described by the (orthonormal) basis vectors $|u_1>$ and $|u_2>$. Then the state vectors

$$|X^{(1)}>=\frac{1}{\sqrt{2}}[|u_1>+|u_2>], \tag{3.48a}$$

$$|X^{(2)}>=\frac{1}{\sqrt{2}}[|u_1>-|u_2>], \tag{3.48b}$$

refer to pure states of the system. According to Eq. (3.56), the density matrices corresponding to these states are

$$\rho^{(1)}=\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix} \tag{3.57a}$$

and

$$\rho^{(2)}=\begin{pmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{pmatrix} \tag{3.57b}$$

which satisfy Eq. (3.55). The density matrix

$$\rho=\begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{pmatrix} \tag{3.58}$$

then represents a mixed state (with $w_1=w_2=\frac{1}{2}$). The average values of an observable $\hat{A}$ represented by the matrix $A=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$, in these three states, as per formula (3.50 [2,3]), are:

$$<\hat{A}>_1=+1,$$

$$<\hat{A}>_2=-1,$$

$$<\overline{\hat{A}}>=0.$$

If $\rho^{(1)}$ and $\rho^{(2)}$ represent *polarized* states, then ρ represents an unpolarized state.

REFERENCE

1. Roman, P., *Advanced Quantum Theory* (Addison-Wesley, Massachusetts 1962) Sections 1.1 to 1.4a.

CHAPTER 4

QUANTUM DYNAMICS

4.1 THE EQUATIONS OF MOTION

The problems of dynamics are, firstly, to determine the variables or parameters, that specify the state of a physical system and, secondly, to describe the evolution of the system in time. In classical mechanics, the dynamical state of a physical system is defined by the values of a set of dynamical variables such as the position co-ordinates and the velocities. The *equations of motion*, which describe the evolution of the physical system in time (temporal development of the system) are, therefore, differential equations in these variables. The assumption that the state of a system at any time t is completely determined, through the equations of motion, by the state of the system at an initial time t_o, is referred to as the *dynamical postulate*. This postulate implies that the equations of motion are *first order* in time[1].

---

1. In classical mechanics, the equations are second order in time for the co-ordinates. This is because the velocities, which are to be treated as independent variables, themselves are first derivatives, with respect to time, of the co-ordinates. The equations of motion are:

   Newtonian: $$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{F};$$

   Lagrange's: $$\frac{\partial L}{\partial q_i} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = 0;$$

   Hamilton's: $$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \text{ and } \frac{dA}{dt} = \frac{\partial A}{\partial t} + \{A, H\};$$

   Hamilton-Jacobi : $$\frac{\partial S}{\partial t} + H = 0,$$

   where, q_i and p_i are the generalised co-ordinates and momenta, and L, H and S are, respectively, the Lagrangian, the Hamiltonian and the Action associated with the system. A is a general dynamical variable which is a function of q_i, p_i and t. $\{A, H\}$ is the *Poisson bracket* defined by

   $$\{A, H\} = \sum_i \left(\frac{\partial A}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial q_i}\frac{\partial A}{\partial p_i}\right).$$

In quantum mechanics, the state of a physical system is represented by a vector defined in an abstract Hilbert space (Postulate II, Chapter 3). Therefore, the equation of motion for a quantum mechanical system could be a differential equation for the state-vector. However, the observable quantities are not the state-vectors, but the *expectation values* of a set of Hermitian operators corresponding to the dynamical variables. The equations of motion in quantum mechanics should be, therefore, concerned with the evolution in time of these expectation values. Now, the expectation value of an operator $\hat{A}$ in the state represented by the (normalized) state-vector ψ, is given by (see Eq. (3.4)),

$$<\hat{A}> \equiv <\hat{A}>_{\psi} = <\psi|\hat{A}|\psi>$$

The variation with time of $<\hat{A}>$ can, therefore, be viewed as arising in one of the following ways:

(a) The state vector ψ changes with time, but $\hat{A}$ remains unchanged.

(b) $\hat{A}$ changes with time, ψ remaining constant.

(c) Both $\hat{A}$ and ψ change with time.

Correspondingly, we have the Schrödinger, the Heisenberg and the Interaction, picture of time development[2]. Of course, the variation with time of $<\hat{A}>$ calculated in any of these pictures should agree with the observed rate of variation of $<\hat{A}>$.

4.1A The Schrödinger Picture

In this case, ψ is a function of t while $\hat{A}$ is not : $\psi \equiv \psi(t)$. The equation of motion is, then, an equation for ψ. The dynamical postulate is that $\psi(t)$ at any time t is completely determined by $\psi(t_0)$ at a given initial time t_0. In view of the fact that ψ is a vector in a linear vector space, the relationship between $\psi(t_0)$ and $\psi(t)$ should be described by a linear operator $\hat{U}(t, t_0)$. The linearity of $\hat{U}$ ensures that the principle of superposition is preserved during the dynamical development of the system. That is, if $\psi(t_0)$ is a certain linear combination of, say, vectors $\phi_i(t_0)$, then $\psi(t)$ is the same linear combination of $\phi_i(t)$. The dynamical postulate could be, thus, restated as follows:

Corresponding to every quantum mechanical system, there exists a family of linear operators $\hat{U}(t, t_0)$, defined on the infinite-dimensional Hilbert space of the system, which describes the evolution of the state-vector from time t_0 to time t:

$$\psi(t) = \hat{U}(t, t_0)\psi(t_0) \tag{4.1}$$

---

2. These different pictures of time-development could be likened to the different ways of describing the rotation of a body relative to a co-ordinate frame. The rotation could be viewed as a rotation of the co-ordinate frame with the body fixed, a rotation of the body with the co-ordinate frame fixed or a combination of the two. The final relative orientation of body and co-ordinate frame should be the same in all the three descriptions.

$\hat{U}$ is called the *evolution* (or *time-development*) *operator*. From the property (2.41a) of a linear operator, it follows that, if

$$\psi(t_0) = \sum_i a_i \phi_i(t_0),$$

then,

$$\psi(t) = \sum_i a_i \hat{U}(t, t_0)\phi(t_0)$$

$$= \sum_i a_i \phi_i(t), \tag{4.2}$$

which proves our earlier statement regarding preservation of the principle of superposition.

The dynamical problem is now reduced to the problem of determining $\hat{U}(t, t_0)$, and we will now address ourselves to this latter problem.

Now, the probability interpretation requires that ψ be normalized to unity at each instant of time (see Postulate III, Chapter 3). Hence,

$$(\psi(t), \psi(t)) = (\psi(t_0), \psi(t_0)) = 1 \tag{4.3}$$

i.e., $$(\hat{U}(t, t_0)\psi(t_0), \hat{U}(t, t_0)\psi(t_0)) = (\psi(t_0), \psi(t_0)),$$

or, by Eq. (2.54), $(\psi(t_0), \hat{U}^+(t, t_0)\hat{U}(t, t_0)\psi(t_0)) = (\psi(t_0), \psi(t_0))$,

so that,

$$\hat{U}^+(t, t_0)\hat{U}(t, t_0) = \hat{1}. \tag{4.4a}$$

Since $\hat{U}$ is defined on an infinite-dimensional Hilbert space, we cannot conclude from Eq. (4.4a) that $\hat{U}$ is unitary (see, *Unitary operators*, Section 2.1). To draw such a conclusion, we should have also the relationship,

$$\hat{U}(t, t_0)\hat{U}^+(t, t_0) = \hat{1}. \tag{4.4b}$$

In fact, $\hat{U}$ satisfies Eq. (4.4b) also, as shown below :

Putting $t = t_0$ in Eq. (4.1), we get,

$$\hat{U}(t_0, t_0) = \hat{1}. \tag{4.5}$$

Substituting t for t_0 and t_1 for t in (4.1), we have,

$$\psi(t_1) = \hat{U}(t_1, t)\psi(t)$$

$$= \hat{U}(t_1, t)\hat{U}(t, t_0)\psi(t_0) \tag{4.6}$$

But,

$$\psi(t_1) = \hat{U}(t_1, t_0)\psi(t_0), \tag{4.1^1}$$

so that,

$$\hat{U}(t_1, t)\hat{U}(t, t_0) = \hat{U}(t_1, t_0) \tag{4.7}$$

Putting $t_1 = t_0$ in (4.7) and using (4.5), we have,

$$\hat{U}(t_0, t)\hat{U}(t, t_0) = \hat{1},$$

or

$$\hat{U}(t,\ t_0)\hat{U}(t_0,\ t) = \hat{1}. \tag{4.8}$$

Multiplying Eq. (4.8) from the left by $\hat{U}(t,\ t_0)\hat{U}^{+}(t,\ t_0)$ and using Eqs. (4.4a) and (4.8), we get Eq. (4.4b). Hence, $\hat{U}$ *is unitary* so that

$$\hat{U}^{+}(t,\ t_0) = \hat{U}^{-1}(t,\ t_0) \tag{4.4}$$

Also, from Eqs. (4.4b), (4.8), and (4.4), we see that

$$\hat{U}^{-1}(t, t_0) = \hat{U}(t_0, t) \tag{4.9}$$

Writing Eq. (4.7) as

$$\hat{U}(t,\ t_1)\ \hat{U}(t_1,\ t_0) = \hat{U}(t,\ t_0),$$

and putting, $t_1 = t - \delta t$, where, δt is infinitesimal, we have,

$$\hat{U}(t, t_0) = \hat{U}(t, t - \delta t)\hat{U}(t - \delta t, t_0), \tag{4.10}$$

Now, $\hat{U}(t, t - \delta t)$ is an infinitesimal unitary operator and could, therefore, be written as (see Eq. (2.63)),

$$\hat{U}(t, t - \delta t) = \hat{1} - (i/\hbar)\delta t\hat{H}(t), \tag{4.11}$$

where, $\hat{H}(t)$ is a Hermitian operator, and the constant $\hbar$ is introduced for convenience of interpretation (see Eq. (4.15b) below). $\hat{H}(t)$ is called the *generator* of the infinitesimal unitary transformation, $\psi(t - \delta t) \rightarrow \psi(t)$.

Substituting (4.11) in (4.10), we get,

$$\hat{U}(t,\ t_0) = \hat{U}(t - \delta t,\ t_0) - (i/\hbar)\delta t \cdot \hat{H}(t)\hat{U}(t - \delta t,\ t_0),$$

or,

$$\frac{\hat{U}(t,\ t_0) - \hat{u}(t - \delta t,\ t_0)}{\delta t} = -(i/\hbar)\hat{H}(t)\hat{U}(t - \delta t,\ t_0).$$

Taking the limit $\delta t \rightarrow 0$ of this equation, we get,

$$\frac{\partial \hat{U}(t, t_0)}{\partial t} = -(i/\hbar)\hat{H}(t)\hat{U}(t, t_0),$$

or,

$$i\hbar\frac{\partial \hat{U}(t, t_0)}{\partial t} = \hat{H}(t)\hat{U}(t, t_0) \tag{4.12}$$

Integrating Eq. (4.12) w.r.t. t, between limits t_0 and t, we get,

$$\int_{t_0}^{t} d\hat{u}(t',\ t_0) = -(i/\hbar)\int_{t_0}^{t} \hat{H}(t')\ \hat{u}(t',\ t_0)\ dt'$$

i.e.,

$$\hat{u}(t,\ t_0) = \hat{1} - \left(\frac{i}{\hbar}\right)\int_{t_0}^{t} \hat{H}(t')\ \hat{u}(t',\ t_0)\ dt' \tag{4.13}$$

Where use has been made of Eq. (4.5).

Equation (4.12) is the differential, and Eq (4.13) the integral, equation for $\hat{U}$. These equations enable us to obtain $\hat{U}(t, t_0)$ for any value of t, from a knowledge of the operator $\hat{H}(t)$.

Applying the operator equation (4.12) to the state-vector $\psi(t_0)$, we get,

$$i\hbar \frac{\partial \hat{U}(t, t_0)\psi(t_0)}{\partial t} = \hat{H}(t)\hat{U}(t, t_0)\psi(t_0).$$

i.e.,

$$i\hbar \frac{\partial \psi(t)}{\partial t} = \hat{H}(t)\psi(t). \tag{4.14}$$

This is the equation of motion for the state-vector, and is known as the *time-dependent Schrödinger equation* (see Eq. (1.16) and the remarks following it).

We, thus, see that the evolution in time of the state-vector ψ could be viewed as the continuous unfolding of a unitary transformation. In analogy with the classical case, where the generator of the infinitesimal canonical transformation corresponding to the temporal development of a mechanical system is the Hamiltonian function, the generator $\hat{H}(t)$ of the unitary transformation, is called the *Hamiltonian* of the system. Just as in the classical case, $\hat{H}$ corresponds to the total energy of the system. If the system has got a classical analogue, then, $\hat{H}(t)$ can be obtained from the classical Hamiltonian of the system in accordance with Postulates IV and V of Chapter 3. If there is no classical analogue, such as in the case of systems with spin, isospin, etc., then, one must rely on intuition or some other circumstantial factors such as symmetry in order to infer the correct form of $\hat{H}(t)$.

Now, Eq. (4.14) can be written as,

$$\frac{\partial}{\partial t}\left(\frac{\hbar}{i} \ln \psi\right) + \hat{H}(t) = 0, \tag{4.14a}$$

from which we see that the equation is the analogue of the Hamilton-Jacobi equation in classical mechanics (see, footnote 1 of this Chapter), with the action S given by

$$S = \frac{\hbar}{i} \ln \psi, \tag{4.15a}$$

or,

$$\psi = \exp\left[(i/\hbar)S\right]. \qquad [\text{cf. Eq. (1.13c)}] \tag{4.15b}$$

We see that the factor $(1/\hbar)$ in Eq. (4.11) is required to make ψ dimensionless, whereas the factor i in that equation, needed to make $\hat{U}$ unitary, makes ψ a complex function.

In the co-ordinate representation, Eq. (4.14) would read [see Eqs. (2.128a), (2.141a), (3.18)]

$$i\hbar \frac{\partial \psi(\mathbf{r}, t)}{\partial t} = \hat{H}(\mathbf{r}, \nabla, t)\psi(\mathbf{r}, t) \tag{4.14b}$$

Problem 4.1: If

$$\hat{H}(\mathbf{r}, \nabla, t) = -\frac{\hbar^2}{2\mu}\nabla^2 + V(\mathbf{r}, t),$$

show that Eq. (4.14b) satisfies the *equation of continuity*,

$$\frac{\partial \rho}{\partial t} + \text{div}\, \mathbf{j} = 0,$$

where the *probability density* ρ and the *probability current density* $\mathbf{j}$ are given by

$$\rho = |\psi|^2, \quad \mathbf{j} = (\hbar/2\mu i)[\psi^*(\nabla\psi) - (\nabla\psi^*)\psi].$$

Problem 4.2: Obtain the conditions on (a) $\hat{H}$ and (b) ψ for the conservation of the norm of ψ.

Often, $\hat{H}$ does not depend explicitly on time. In that case, we have, from Eq. (4.12),

$$\int_{t_0}^{t} \frac{d\hat{U}(t', t_0)}{\hat{U}(t', t_0)} = -(i/\hbar)\hat{H}\int_{t_0}^{t} dt',$$

or,
$$\hat{U}(t, t_0) = e^{-(i/\hbar)\hat{H}(t - t_0)}, \tag{4.16}$$

so that,

$$\psi(\mathbf{r}, t) = e^{-(i/\hbar)\hat{H}(t - t_0)}\psi(\mathbf{r}, t_0) \tag{4.17a}$$

Choosing $t_0 = 0$, and writing $\psi(\mathbf{r}, 0) \equiv \Phi(\mathbf{r})$,

we have,

$$\psi(\mathbf{r}, t) = e^{-(i/\hbar)\hat{H}t}\Phi(\mathbf{r}). \tag{4.17b}$$

Now, Φ can be expanded in terms of the complete set of eigenvectors of the Hermitian operator $\hat{H}$. Let the eigenvectors be $\phi_i(\mathbf{r})$:

$$\hat{H}\phi_i(\mathbf{r}) = E_i\phi_i(\mathbf{r}) \tag{4.18}$$

Then,
$$\Phi(\mathbf{r}) = \sum_i c_i\phi_i(\mathbf{r}) \tag{4.19}$$

Substituting (4.19) in (4.17b), we get,

$$\psi(\mathbf{r}, t) = \sum_i c_i\psi_i(\mathbf{r}, t), \tag{4.20}$$

where,

$$\psi_i(\mathbf{r}, t) = \phi_i(\mathbf{r})e^{-(i/\hbar)E_i t}$$

$$= \phi_i(\mathbf{r})\chi(t). \tag{4.21}$$

Obviously,

$$\hat{H}\psi_i(\mathbf{r}, t) = E_i\psi_i(\mathbf{r}, t), \tag{4.22}$$

so that, ψ_i is also an eigenvector of $\hat{H}$ belonging to the eigenvalue E_i.

Equation (4.18) is referred to as the *time-independent Schrödinger equation.* Each ϕ_i represents a *stationary state* of the system with energy E_i, which evolves in time according to $e^{-(i/\hbar)E_i t}$:

$$\hat{U}(t,0)\phi_i(\mathbf{r}) = e^{-(i/\hbar)\hat{H}t}\phi_i(\mathbf{r})$$

$$= e^{-(i/\hbar)E_i t}\phi_i(\mathbf{r}). \tag{4.23}$$

The name, 'stationary state' arises from the following :

(i) A system which is initially in the state ϕ_i continues to be in this state since the evolution operator does not mix different ϕ_i's (see Eq. (4.2)).

(ii) The expectation values of operators, and, in particular, the energy, in the state does not change with time (shown below).

Now,

$$\frac{d}{dt}\langle\hat{A}\rangle = \frac{d}{dt}\langle\psi(t)\,|\,\hat{A}\,|\,\psi(t)\rangle$$

$$= \langle\frac{\partial\psi}{\partial t}|\hat{A}|\psi\rangle + \langle\psi|\hat{A}|\frac{\partial\psi}{\partial t}\rangle \tag{4.24}$$

But, from Eq. (4.14),

$$\frac{\partial\psi}{\partial t} = -(i/\hbar)\hat{H}\psi, \tag{4.14'}$$

so that,

$$\frac{d}{dt}\langle\hat{A}\rangle = \langle(-i/\hbar)\hat{H}\psi\,|\,\hat{A}\,|\,\psi\rangle + \langle\psi\,|\,\hat{A}\,|\,(-i/\hbar)\hat{H}\psi\rangle$$

$$= (i/\hbar)\langle\psi\,|\,(\hat{H}\hat{A} - \hat{A}\hat{H})\,|\,\psi\rangle$$

$$= \langle\frac{1}{i\hbar}[\hat{A},\hat{H}]\rangle, \tag{4.25}$$

where $[\hat{A},\hat{H}]$ is the *commutator,* or *commutator bracket,* of $\hat{A}$ and $\hat{H}$.

If the system is in one of the stationary states, say

$$\psi = \psi_i \equiv \phi_i(\mathbf{r})e^{-(i/\hbar)E_i t}, \text{ then,}$$

$$<\psi_i\,|\,(\hat{A}\hat{H} - \hat{H}\hat{A})\,|\,\psi_i> \;=\; <\phi_i\,|\,(\hat{A}\hat{H} - \hat{H}\hat{A})\,|\,\phi_i>$$

$$= (E_i - E_i^*)<\phi_i\,|\,\hat{A}\,|\,\phi_i> \;=\; 0, \tag{4.26}$$

where, use has been made of Eq. (4.18) and of the fact that eigenvalues of a Hermitian operator are real (so that, $E_i^* = E_i$). Thus, $\frac{d\langle\hat{A}\rangle}{dt} = 0$ for the stationary state. $\frac{d\langle\hat{A}\rangle}{dt} = 0$ also when $[\hat{A},\hat{H}] = \hat{0}$. This case will be further discussed in Chapter 6.

In the presence of an external field, $\hat{A}$ may depend explicitly on time. In that case, the term $\langle\psi \left| \frac{\partial \hat{A}}{\partial t} \right| \psi\rangle$ should be added to the R.H.S. of Eq. (4.24). Then, in place of (4.25), we have,

$$\frac{d\langle\hat{A}\rangle}{dt} = \langle\frac{\partial \hat{A}}{\partial t}\rangle + \langle(1/i\hbar)[\hat{A}, \hat{H}]\rangle. \tag{4.25a}$$

Comparing Eq. (4.25a) with the equation of motion for the dynamical variables A in classical mechanics (see footnote of this chapter), we see that *the expectation values of operators obey the same equation of motion in quantum mechanics as the dynamical variables in classical mechanics*[3], provided we identify the commutator bracket divided by $(i\hbar)$ with the 'quantum mechanical Poisson bracket'. The statement in italics, is known as *Ehrenfest's theorem.*

The identification of the commutator bracket divided by $i\hbar$, with the Poisson bracket is, in fact, suggested also by the identity of the algebra the two brackets obey, exhibited in Table 4.1. In this table, we also give certain examples illustrating the similarity of the two brackets. It is reasonable to conclude from this that Poisson bracket is the classical limit of the expectation value of the commutator bracket divided by $i\hbar$.

$$< \frac{1}{i\hbar}[\hat{A}, \hat{B}] > \xrightarrow[\text{classical}]{} \{A, B\}. \tag{4.27}$$

Problem 4.3: Establish the following relationships in the case of the linear harmonic oscillator :

$$\text{(a)} \quad \frac{d}{dt} < \hat{x} > = \frac{< \hat{p} >}{m}$$

$$\text{(b)} \quad \frac{d < \hat{p} >}{dt} = < -\frac{d\hat{V}}{dx} >$$

4.1B The Heisenberg Picture

In order to distinguish the state-vectors and operators in this picture from those of the Schrödinger picture, we will use the subscript 'H'. Thus ψ_H and $\hat{A}_H$, respectively, denote a statevector and an operator in the Heisenberg picture.

ψ_H is time-independent, but $\hat{A}_H$ depends on time. We note that if we define $\psi_H(t)$ by

$$\psi_H(t) = \hat{U}^{-1}(t, t_0)\psi(t), \tag{4.28a}$$

where, $\psi(t)$ is the state vector in the Schrödinger picture, then, ψ_H is independent of time. For, from Eq. (4.1), we have,

---

3. Note that $\langle\frac{\partial \hat{A}}{\partial t}\rangle \equiv \frac{\partial}{\partial t} < \hat{A} >$.

$$\psi_H(t) = \hat{U}^{-1}(t, t_0)\psi(t) = \hat{U}^{-1}(t, t_0)\hat{U}(t, t_0)\psi(t_0) = \psi(t_0) \tag{4.28b}$$

Table 4.1 Comparison of Commutator and Poisson Brackets

Commutator Bracket [,]	Poisson bracket { , }
$[\hat{A}, \hat{B}] = -[\hat{B}, \hat{A}]$	$\{A, B\} = -\{B, A\}$
$[\hat{A}, c] = \hat{0}$,	$\{A, c\} = 0$
$[(\hat{A} + \hat{B}), \hat{C}] = [\hat{A}, \hat{C}] + [\hat{B}, \hat{C}]$	$\{(A + B), C\} = \{A, C\} + \{B, C\}$
$[\hat{A}\hat{B}, \hat{C}] = [\hat{A}, \hat{C}]\hat{B} + \hat{A}[\hat{B}, \hat{C}]$	$\{AB, C\} = \{A, C\}B + A\{B, C\}$
$\frac{1}{i\hbar}[\hat{q}_i, \hat{p}_j] = \hat{\delta}_{ij}$	$\{q_i, p_j\} = \delta_{ij}$
$\frac{1}{i\hbar}[\hat{L}_x, \hat{L}_y] = \hat{L}_z$	$\{L_x, L_y\} = L_z$

c--a scalar number.

q_i, p_j—generalized co-ordinates and momenta.

L_x, L_y, L_z—components of angular momentum.

Since $\hat{U}(t, t_0)$ is unitary, Eq. (4.28a) represents a unitary transformation (a change of basis) in the vector space. Therefore, $\hat{A}_H$ is related to the corresponding operator $\hat{A}$ in the Schrödinger picture by (see Eq. (2.126)),

$$\begin{aligned} \hat{A}_H(t) &= \hat{U}^{-1}(t,\ t_0)\hat{A}\hat{U}(t,\ t_0) \\ &= U^{\dagger}(t,\ t_0)\hat{A}\hat{U}(t,\ t_0). \end{aligned} \tag{4.29}$$

Thus,

$$\frac{d\hat{A}_H}{dt} = \frac{\partial U^{\dagger}}{\partial t}\hat{A}\hat{U} + \hat{U}^{\dagger}\hat{A}\frac{\partial \hat{U}}{\partial t}. \tag{4.30}$$

But, from Eq. (4.12),

$$\frac{\partial \hat{U}}{\partial t} = -(i/\hbar)\hat{H}\hat{U}, \tag{4.31a}$$

and

$$\frac{\partial U^{\dagger}}{\partial t} = -(i/\hbar)\hat{U}^{\dagger}\hat{H}, \tag{4.31b}$$

so that,

$$\begin{aligned} \frac{d\hat{A}_H}{dt} &= (i/\hbar)\{\hat{U}^{\dagger}\hat{H}\hat{A}\hat{U} - \hat{U}^{\dagger}\hat{A}\hat{H}\hat{U}\} \\ &= (i/\hbar)\{U^{\dagger}\hat{H}\hat{U}\hat{U}^{\dagger}\hat{A}\hat{U} - \hat{U}^{\dagger}\hat{A}\hat{U}\hat{U}^{\dagger}\hat{H}\hat{U}\} \\ &= (i/\hbar)(\hat{H}_H\hat{A}_H - \hat{A}_H\hat{H}_H) \\ &= \frac{1}{i\hbar}[\hat{A}_H, \hat{H}_H]. \end{aligned} \tag{4.32a}$$

If $\hat{A}$ depends explicitly on time, then,

$$\frac{d\hat{A}_H}{dt} = \frac{\partial \hat{A}_H}{\partial t} + \frac{1}{i\hbar}[\hat{A}_H, \hat{H}_H]. \tag{4.32b}$$

Equations (4.32a, b) are called the *Heisenberg's equations of motion* for the operator $\hat{A}_H$. They are identical in form to the Hamilton's equation of motion for a dynamical variable in classical mechanics (see footnote 1), the only difference being that the place of the Poisson bracket in the latter is taken by the commutator bracket upon $i\hbar$ in the former.

For the basic canonical operators $\hat{q}_j$ and $\hat{p}_j$, we have, from Eq. (4.32a),

$$\frac{d\hat{q}_i}{dt} = \frac{1}{i\hbar}[\hat{q}_j, \hat{H}]$$

$$= \frac{1}{i\hbar}\left(i\hbar\frac{\partial \hat{H}}{\partial \hat{p}_j}\right) = \frac{\partial \hat{H}}{\partial \hat{p}_j}, \text{ by Eq. (3.14c) ;} \tag{4.33a}$$

$$\frac{dp_j}{dt} = \frac{1}{i\hbar}[\hat{p}_j, \hat{H}]$$

$$= \frac{1}{i\hbar}\left(-i\hbar\frac{\partial \hat{H}}{\partial \hat{q}_j}\right) = -\frac{\partial \hat{H}}{\partial \hat{q}_j}, \text{ by Eq. (3.14d).} \tag{4.33b}$$

Again, these equations are identical with the corresponding canonical equations of Hamilton in classical mechanics.

The equation of motion for the expectation value $< \hat{A}_H >$ is given by

$$\frac{d}{dt}\langle \hat{A}_H \rangle \equiv \frac{d}{dt}\langle \psi_H \mid \hat{A}_H(t) \mid \psi_H \rangle$$

$$= \langle \psi_H \mid \frac{d\hat{A}_H}{dt} \mid \psi_H \rangle$$

$$= \langle \frac{1}{i\hbar}[\hat{A}_H, \hat{H}_H] \rangle, \tag{4.34a}$$

if $\hat{A}_H$ has no explicit dependence on time, and

$$\frac{d\langle \hat{A}_H \rangle}{dt} = \langle \frac{\partial \hat{A}_H}{\partial t} \rangle + \langle \frac{1}{i\hbar}[\hat{A}_H, \hat{H}_H] \rangle, \tag{4.34b}$$

if $\hat{A}_H$ depends explicitly on time.

Thus, as expected, the equations of motion for the expectation values in the Heisenberg picture and the Schrödinger picture are the same.

The basic equation of motion in the Schrödinger picture is Eq. (4.14), whereas in the Heisenberg picture, it is Eq. (4.32b). We, thus, see that the Heisenberg picture emphasises Hamilton's formalism of classical mechanics while the Schrödinger picture emphasises the Hamilton-Jacobi theory. In the former case,

the emphasis is on the physical observables, whereas in the latter case, it is on the function ψ. The methods of solution of the mechanical problem in the two pictures will also reflect this difference in the emphasis, as we will illustrate in Section 4.2.

4.1C The Interaction Picture

In this picture, the advantages of the Heisenberg and the Schrödinger pictures are sought to be combined. Such an approach is useful when the Hamiltonian can be split up into two parts, one part independent of time and the other part dependent on time. Such could be the case, for example, when the system is in an external field. The time-independent part would represent the Hamiltonian of the system in the absence of the external field and the time-dependent part that arising from the presence of the external field (Section 8.4 will discuss such cases).

Let
$$\hat{H}(t) = \hat{H}^{(0)} + \hat{H}^{(1)}(t), \tag{4.35}$$

$\hat{H}^{(0)}$ being independent of time.

According to Eq. (4.16), the time-development operator in the absence of $\hat{H}^{(1)}(t)$ is given by

$$\hat{U}_0(t, t_0) = \exp\left[-(i/\hbar)\hat{H}^{(0)}(t - t_0)\right] \tag{4.36}$$

The state-vector $\psi_I(t)$ and operator $\hat{A}_I(t)$ in the interaction picture are defined by

$$\psi_I(t) = \hat{U}_0^{-1}(t, t_0)\psi(t),$$

$$= \exp\left[(i/\hbar)\hat{H}^{(0)}(t - t_0)\right]\psi(t), \tag{4.37}$$

and

$$\hat{A}_I(t) = \hat{U}_0^{-1}(t, t_0)\hat{A}\hat{U}_0(t, t_0) \tag{4.38}$$

$$= \exp\left[(i/\hbar)\hat{H}^{(0)}(t - t_0)\right]\hat{A}\exp\left[-(i/\hbar)\hat{H}^{(0)}(t - t_0)\right]$$

where, $\psi(t)$ and $\hat{A}$ are the state vector and the operator, respectively, in the Schrödinger picture, so that

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}(t)\psi(t)$$

$$= [\hat{H}^{(0)} + \hat{H}^{(1)}(t)]\psi(t), \tag{4.39a}$$

and

$$\frac{d\hat{A}}{dt} = \frac{\partial\hat{A}}{dt}, \tag{4.39b}$$

From Eqs. (4.37) and (4.39a), we get,

$$i\hbar\frac{\partial\psi_I}{\partial t} = \hat{H}_I^{(1)}\psi_i(t), \tag{4.40}$$

where,
$$\hat{H}_I^{(1)} = \hat{U}_0^{-1}(t,\ t_0)\hat{H}^{(1)}\hat{U}_0(t,\ t_0), \tag{4.40a}$$

while, from Eqs. (4.38) and (4.39b), we have,

$$\frac{d\hat{A}_I}{dt} = \frac{\partial \hat{A} I}{\partial t} + \frac{1}{i\hbar}[\hat{A}_I, \hat{H}_I^{(0)}], \tag{4.41}$$

with

$$\hat{H}_I^{(0)} \equiv \hat{U}_0^{-1}(t, t_0)\hat{H}^{(0)}\hat{U}_0(t, t_0)$$

$$= \hat{H}^{(0)}. \tag{4.41a}$$

We see that the state-vector in the interaction picture is determined by a Schrödinger equation with the Hamiltonian $\hat{H}_I^{(1)}$, while the operators obey the Heisenberg's equation with the Hamiltonian $\hat{H}^{(0)}$.

Problem 4.4: Show that the evolution operator $\hat{U}_I(t,\ t_0)$ defined by the equation,

$$i\hbar \frac{\partial U_I(t,\ t_0)}{\partial t} = \hat{H}_I^{(1)}(t)\hat{U}_I(t,\ t_0),$$

satisfies the relationship,

$$\hat{U}(t,\ t_0) = \hat{U}_0(t,\ t_0)\hat{U}_I(t,\ t_0),$$

where,

$$\hat{U}(t,\ t_0)\psi(t_0) = \psi(t).$$

Hence show that $\psi_I(t) = \hat{U}_I(t,\ t_0)\psi_I(t_n)$.

The equation of motion for the expectation value,

$$< \psi_I \mid \hat{A}_I \mid \psi_I > \equiv < \hat{A}_I > .$$

is easily shown, using (4.41) and (4.40), to be

$$\frac{d}{dt}\langle \hat{A}_I \rangle = \langle \frac{\partial \hat{A}_I}{\partial t} \rangle + \langle \frac{1}{i\hbar}[\hat{A}_I,\ \hat{H}_I] \rangle \tag{4.42}$$

where,

$$\hat{H}_I = \hat{H}_I^{(0)} + \hat{H}_I^{(1)} \tag{4.42a}$$

4.2 ILLUSTRATIVE APPLICATIONS

4.2A The Linear Harmonic Oscillator

We will illustrate the difference between the Schrödinger and the Heisenberg pictures by applying the two methods to the solution of the problem of the linear harmonic oscillator for which the Hamiltonian is given by (see Eq. (3.9)),

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}K\hat{x}^2.$$

Schrödinger Picture

In the Schrödinger picture, $\hat{p}$ and $\hat{x}$ are independent of time. In the co-ordinate representation (Eq. (3.18)),

$$\hat{p} \equiv -i\hbar\frac{d}{dx} \text{ and } \hat{x} \equiv x,$$

so that

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}Kx^2 \tag{4.43}$$

The time-dependent Schrödinger equation (Eq. (4.14)) is

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}Kx^2\right)\psi(x,t). \tag{4.44}$$

Since $\hat{H}$ is independent of time, $\psi(x,t)$ can be written, choosing $t_0 = 0$, as (see Eqs. (4.17b) and (4.18)),

$$\psi(x,t) = \exp[-(i/\hbar)\hat{H}t]u(x), \tag{4.45}$$

with

$$\hat{H}u_n(x) = E_n u_n(x) \tag{4.46}$$

That is

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}Kx^2\right)u_n(x) = E_n u_n(x) \tag{4.46a}$$

which is the time-independent Schrödinger equation for the linear-harmonic oscillator. From Eqs. (4.45) and (4.46), we get,

$$\psi_n(x,t) = \exp[-(i/\hbar)E_n t]u_n(x)$$

$$= \chi_n(t)u_n(x). \tag{4.45a}$$

Thus, what is involved in the reduction of Eq. (4.44) to the form (4.46a) is the technique of *separation of variables*.

Now, Eq. (4.46) is an eigenvalue equation for the Hamiltonian. Therefore, the values of E_n, permitted by the equation, are the energies that the harmonic oscillator can have. The state of the oscillator with energy E_n is represented by the state-vector ('wave function') $u_n(x)$. The problem of determining the E_n and the u_n is reduced to solving the differential equation (4.46a).

Now, Eq. (4.46a) resembles Eq. (E.9); in fact, the former could be made to look identical with the latter with the substitution,

$$\zeta = \alpha x;\ \alpha = \left(\frac{mK}{\hbar^2}\right)^{1/4} = \sqrt{\frac{m\omega}{\hbar}};\ (1+2n) = \frac{2E_n}{\hbar\omega}, \tag{4.47a}$$

where,

$$\omega = \sqrt{\frac{K}{m}}, \tag{4.47b}$$

is the classical frequency of the oscillator.

Then, equation (4.46a) reduces to

$$\frac{d^2\phi_n}{d\zeta^2} + (1 + 2n - \zeta^2)\phi_n = 0. \tag{4.48}$$

where,

$$\phi_n(\zeta) \propto u_n(x). \tag{4.49}$$

From Eq. (E. 8),

$$\phi_n(\zeta) = \frac{1}{[\sqrt{\pi}\, 2^n n!]^{1/2}} \exp(-\zeta^2/2) H_n(\zeta), \tag{4.50a}$$

or,

$$u_n(x) = \frac{\sqrt{\alpha}}{[\sqrt{\pi}\, 2^n n!]^{1/2}} \exp\left(-\frac{1}{2}\alpha^2 x^2\right) H_n(\alpha x), \tag{4.50b}$$

which are the eigenfunctions. The corresponding energy eigenvalues are, from Eq. (4.47a),

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega, \qquad n = 0, 1, 2, \ldots \tag{4.51}$$

We see that Eq. (4.51) differs from the corresponding classical formula[4] in two important aspects:

(i) The energy levels are discrete and equispaced. Discreteness is a property of bound systems, but the equispacedness is a characteristic of the oscillator.

(ii) The lowest (referred to as 'ground-state') energy is not zero, but is equal to $\frac{1}{2}\hbar\omega$. This *zero-point energy* could be attributed to the uncertainty relationship [see Section 3.2, Eq. (3.33b)]. The energy levels are shown in Fig. 4.1.

From Eq. (E.7) and (4.50a), we have,

$$\phi_n(-\zeta) = (-1)^n \phi_n(\zeta). \tag{4.52}$$

Thus, $\phi_n(\zeta)$ is an even or an odd function of ζ according as n is even or odd. The operation which transforms x to $-x$, is referred to as the *Parity operation* (see Section 6.2D), and the behaviour of a function under the parity operation determines the *parity* of the function. A function $f(x)$ has even parity if $f(-x) = f(x)$ and odd parity if $f(-x) = -f(x)$, whereas it has no definite parity if $f(-x) \neq \pm f(x)$. Invariance of the Hamiltonian under the parity operation (which represents space

4. The classical formula corresponding to (4.51) is $E = J\nu$, where ν is the linear frequency and J is the action variable (see H. Goldstein, *Classical Mechanics* (Addison-Wesley, Massachusetts, 1961), p. 294). Thus, Eq. (4.51) implies quantization of the action according to the formula, $J = \left(n + \frac{1}{2}\right)h$.

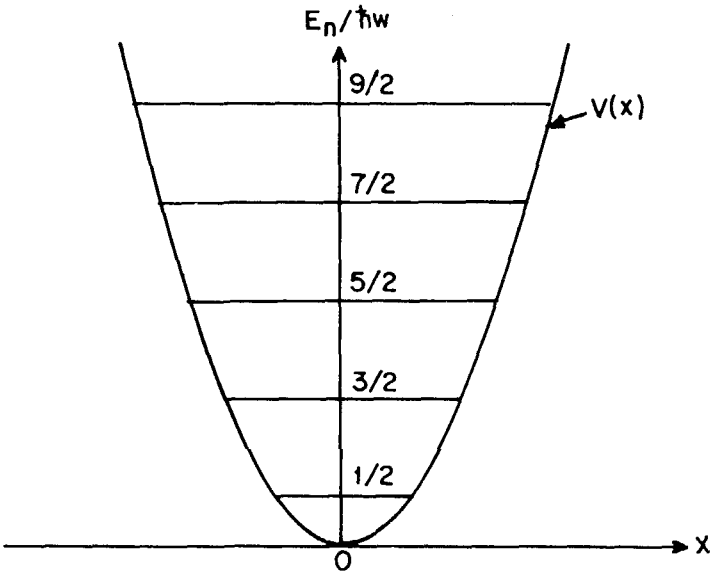


Fig. 4.1. The potential and the energy levels of the linear harmonic oscillator.

inversion, or the changing of a co-ordinate system from a right-handed one to a left-handed one) requires that wave functions of physical systems be of definite parity. Eq. (4.52) shows that this requirement is met by the wave functions of the linear harmonic oscillator. It also shows that the parity of the wave function ϕ_n is even or odd according as the *oscillator quantum number n* is even or odd.

Now, according to Eq. (4.52), $\phi_n(-\infty) = \phi_n(+\infty)$ for n even and $\phi_n(-\infty) = -\phi_n(+\infty)$ for n odd. Thus, $\phi_n(\zeta)$ approaches the ζ-axis from the same direction for even n, and from the opposite directions for odd n. Hence the number of zeroes of $\phi_n(\zeta)$, excluding the ones at $\zeta = \pm\infty$, is even or odd according as n is even or odd. In fact, the number of zeroes (these are referred to as *nodes* of ϕ_n) is exactly equal to n (see Fig. E.1).

We note the following important differences with respect to a classical oscillator:

(i) The amplitude of oscillation of the classical oscillator in the mode n is given, according to Eq. (4.48), by

$$\zeta_n = \sqrt{1+2n}, \tag{4.53}$$

so that the probability of finding the particle outside this range is zero. This amplitude is also shown in Fig. E.1 (the shaded region). We see that the wave function ϕ_n (the probability is proportional to $|\phi_n|^2$) does not go to zero at ζ_n, even though it goes to zero rapidly outside this range. This is a general feature of quantum mechanical wave functions: the wave function

goes to zero at a classical boundary only if the potential at the boundary is infinite.[5] In other words, a perfectly 'opaque' boundary is represented, in quantum mechanics, by an infinite potential.

(ii) The probability distribution, $P_n(\zeta) \equiv |\phi_n|^2$, exhibits maxima and minima within the classical interval $-\zeta_n \le \zeta \le \zeta_n$ (shown in Fig. 4.2, for $n = 2$). This is in contrast with the smoothly varying classical probability distribution $P_n^{cl}(\zeta)$, which is given by[6],

$$P_n^{cl}(\zeta) = \frac{1}{\pi\sqrt{1+2n-\zeta^2}}. \tag{4.54}$$

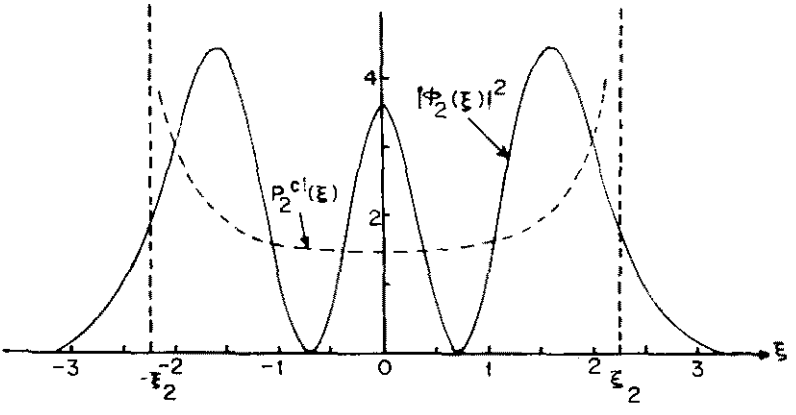


Fig. 4.2. The classical (broken curve) and the quantum mechanical (solid curve) position probability distributions corresponding to oscillator quantum number $n = 2$.

5. Note that ϕ_n will go to zero at ζ_n, if α (that is, K; see Eq. (4.47a)) is set equal to infinity at that point.

6. The classical probability $P_n^{(cl)}(\zeta)d\zeta$ for the particle to be between ζ and $\zeta + d\zeta$ could be defined as the ratio (dt'/τ), where dt' is the time the oscillator spends between ζ and $\zeta + d\zeta$ during the period τ. Now, since the amplitude of oscillation is given by Eq. (4.53), we have,

$$\zeta(t) = \zeta_n \sin \omega t,$$

and
$$\frac{d\zeta}{dt} = \omega\zeta_n \cos \omega t = \omega\sqrt{\zeta_n^2 - \zeta^2}$$

Thus,
$$\frac{dt'}{\tau} = \frac{2dt}{\tau} = \frac{1}{\tau} \cdot \frac{2d\zeta}{\omega(\zeta_n^2 - \zeta^2)^{1/2}}$$

But $\tau = 2\pi/\omega$, so that,

$$P_n^{cl}(\zeta)d\zeta = \frac{2dt}{\tau} = \frac{d\zeta}{\pi\sqrt{\zeta_n^2 - \zeta^2}}$$

From Eqs. (4.45a) and (4.51) we have,

$$\psi_n(x,t) = u_n(x)e^{-i\omega_n t}, \tag{4.55}$$

where, $$\omega_n = \left(n + \frac{1}{2}\right)\omega. \tag{4.51'}$$

The expectation value of any operator could be calculated using the wave function $\psi_n(x,t)$. An arbitrary wave function $\Psi(x,t)$ would be a linear superposition of the ψ_n:

$$\Psi(x,t) = \sum_n c_n \psi_n(x,t). \tag{4.56}$$

Problem 4.5 : Show that, in the state ϕ_n, the uncertainty product $\Delta x \cdot \Delta p$ is given by,

$$\Delta x \cdot \Delta p = \left(n + \frac{1}{2}\right)\hbar.$$

Problem 4.6: Show that, in the case of the 3-dimensional, isotropic harmonic oscillator, the energy levels are given by

$$E_N = \left(N + \frac{3}{2}\right)\hbar\omega,$$

with N = 0, 1, 2, ...

Heisenberg Picture

The operators are functions of t, so that (we drop the label H attached to Heisenberg operators),

$$\hat{H}(t) = \frac{\{\hat{p}(t)\}^2}{2m} + \frac{1}{2}K\{\hat{x}(t)\}^2$$

$$= \frac{1}{2m}[\hat{p}^2(t) + m^2\omega^2\hat{x}^2(t)] \tag{4.57}$$

The equations of motion for $\hat{x}$ and $\hat{p}$, according to Eqs. (4.33a, b), are:

$$\frac{d\hat{x}}{dt} = \frac{\partial \hat{H}}{\partial \hat{p}} = \frac{\hat{p}}{m}, \tag{4.58a}$$

and $$\frac{d\hat{p}}{dt} = -\frac{\partial H}{\partial \hat{x}} = -m\omega^2\hat{x}. \tag{4.58b}$$

Differentiating (4.58a) w.r.t. t once again, and substituting for $\frac{d\hat{p}}{dt}$ from Eq. (4.58b), we get,

$$\frac{d^2\hat{x}}{dt^2} + \omega^2\hat{x} = \hat{0}. \tag{4.58}$$

which is of the same form as the classical equation of motion for the harmonic oscillator.

The solution of Eq. (4.58) is

$$\hat{x}(t) = \hat{c}_1 e^{-i\omega t} + \hat{c}_2 e^{-i\omega t}$$

$$= \hat{a}_1 \cos \omega t + \hat{a}_2 \sin \omega t \tag{4.59}$$

$$\hat{a}_1 = \hat{x}(0) \equiv \hat{x}_0, \text{ say,}$$

and

$$\hat{a}_2 = \frac{1}{\omega} \frac{d\hat{x}}{dt}\bigg|_{t=0} = \frac{\hat{p}(0)}{m\omega} = \frac{\hat{p}_0}{m\omega},$$

so that,

$$\hat{x}(t) = \hat{x}_0 \cos \omega t + (\hat{p}_0 / m\omega) \sin \omega t, \tag{4.59a}$$

$$\hat{p}(t) = m\frac{d\hat{x}}{dt} = \hat{p}_0 \cos \omega t - m\omega \hat{x}_0 \sin \omega t \tag{4.59b}$$

Substituting in (4.57) from (4.59a, b), we get,

$$\hat{H} = \frac{1}{2m}(\hat{p}_0^2 + m^2\omega^2\hat{x}_0^2), \tag{4.57a}$$

which is independent of time.

To obtain the energy levels of the oscillator, we have to calculate the matrix of $\hat{H}$ and diagonalize it. If we use the $u_n(x)$ given by Eq. (4.50b), which are also state vectors in the Heisenberg picture (see Eq. (4.28b)), H would be diagonal. For,

$$H_{n'n} = \frac{1}{2m}(p_0^2)_{n'n} + \frac{1}{2}m\omega^2(x_0^2)_{n'n}. \tag{4.60a}$$

Now, $\hat{p}_0$ and $\hat{x}_0$ are the Heisenberg operators at time $t = 0$, and are hence identical with the corresponding operators in the Schrödinger picture (Eq. (4.29)). Thus,

$$\hat{x}_0 \equiv x; \quad \hat{p}_0 \equiv -i\hbar\frac{d}{dx}. \tag{4.61}$$

Then,

$$(p_0^2)_{n'n} \equiv (-i\hbar)^2 \int_{-\infty}^{\infty} u_{n'}^*(x) \frac{d^2}{dx^2} u_n(x) dx$$

$$= -\hbar^2\alpha^2 \int_{-\infty}^{\infty} \phi_{n'}^*(\zeta) \phi''_n(\zeta) d\zeta,$$

where ζ is given by Eq. (4.47a). From Eq. (E.9), we have,

$$\phi''_n = [\zeta^2 - (1 + 2n)]\phi_n \tag{4.62}$$

so that,

$$\frac{1}{2m}(p_0^2)_{n'n} = \frac{\hbar^2\alpha^2}{2m}\left[(2n+1)\delta_{nn'} - \alpha^2(x_0^2)_{n'n}\right],$$

where, the orthonormality of the ϕ_n (Eq. (E.10)) has been used.

Now,

$$\frac{\hbar^2\alpha^2}{2m}=\frac{\hbar}{2}\sqrt{\frac{K}{m}}=\frac{1}{2}\hbar\omega;\ \frac{\hbar^2\alpha^4}{2m}=\frac{1}{2}m\omega^2.$$

Hence,

$$\frac{1}{2m}(p_0^2)_{n'n}=\left(n+\frac{1}{2}\right)\hbar\omega\delta_{nn'}-\frac{1}{2}m\omega^2(x_0^2)_{n'n}. \tag{4.63}$$

From Eqs. (4.60a) and (4.63), we get,

$$H_{n'n}=\left(n+\frac{1}{2}\right)\hbar\omega\,\delta_{nn'}. \tag{4.60b}$$

Thus, the eigenvalues of $\hat{H}$ are given by

$$H_{nn}\equiv E_n=\left(n+\frac{1}{2}\right)\hbar\omega,\quad n=0,\,1,\,2\ldots \tag{4.64}$$

which are the same as those given by Eq. (4.51) in the Schrödinger picture.

Using Eq. (E.11a), we get,

$$(x_0)_{n'n}=\frac{1}{\alpha}\left[\sqrt{\frac{n+1}{2}}\delta_{n',n+1}+\sqrt{\frac{n}{2}}\delta_{n',n-1}\right]$$

$$=\sqrt{\frac{\hbar}{2m\omega}}[\sqrt{n}\,\delta_{n,n'+1}+\sqrt{n+1}\,\delta_{n,n'-1}] \tag{4.65}$$

Also,

$$(p_0)_{n'n}=\left(-i\hbar\frac{d}{dx}\right)_{n'n}\equiv-i\hbar\int_{-\infty}^{\infty}u_{n'}(x)\frac{d}{dx}u_n(x)dx$$

$$=-i\hbar\alpha\int_{-\infty}^{\infty}\phi_{n'}(\zeta)\frac{d}{d\zeta}\phi_n(\zeta)d\zeta,$$

From Eqs. (E.8) and (E.6a), we have

$$\frac{d}{d\zeta}\phi_n(\zeta)=-\zeta\phi_n+\sqrt{2n}\,\phi_{n-1}. \tag{4.66}$$

Hence,

$$(p_0)_{n'n}=i\hbar\alpha^2(x_0)_{n'n}-i\hbar\alpha\sqrt{2n}\,\delta_{n,n'-1}$$

$$=i\hbar\alpha\left[\sqrt{\frac{n}{2}}\delta_{n,n'+1}-\sqrt{\frac{n+1}{2}}\delta_{n,n'-1}\right]$$

$$=i\sqrt{\frac{m\omega\hbar}{2}}[\sqrt{n}\,\delta_{n,n'+1}-\sqrt{n+1}\,\delta_{n,n'-1}]. \tag{4.67}$$

From Eqs. (4.59a, b), (4.65) and (4.67), we get,

$$\{x(t)\}_{n'n}=\sqrt{\frac{\hbar}{2m\omega}}[\sqrt{n}\,\exp(-i\omega t)\delta_{n,n'+1}$$

$$+\sqrt{n+1}\,\exp(i\omega t)\delta_{n,n'-1}] \tag{4.68}$$

and
$$\{p(t)\}_{n'n} = i\sqrt{\frac{m\hbar\omega}{2}}[\sqrt{n+1}\ \exp(i\omega t)\delta_{nn'-1} - \sqrt{n}\ \exp(-i\omega t)\delta_{n,n'+1}]$$
$$= m < n' \left| \frac{d\hat{x}}{dt} \right| n >, \tag{4.69}$$

where, $| n > \equiv | u_n >$.

Thus, the expectation values of $\hat{x}$ and $\hat{p}$ are zero:
$$< n \mid \hat{x} \mid n > = 0 = < n \mid \hat{p} \mid n > . \tag{4.70}$$

From Eqs (4.65), (4.67) and (4.60b), we get the matrices x_0, p_0 and H representing $\hat{x}_0$, $\hat{p}_0$ and $\hat{H}$ respectively:

$$(x_0) = \sqrt{\frac{\hbar}{2m\omega}} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & \ldots. \\ \sqrt{1} & 0 & \sqrt{2} & 0 & 0 & \ldots. \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 & \ldots. \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} & \ldots. \end{pmatrix} \tag{4.71}$$

$$(P_0) = i\sqrt{\frac{m\hbar\omega}{2}} \begin{pmatrix} 0 & -\sqrt{1} & 0 & 0 & 0 & .. \\ \sqrt{1} & 0 & -\sqrt{2} & 0 & 0 & .. \\ 0 & \sqrt{2} & 0 & -\sqrt{3} & 0 & .. \\ 0 & 0 & \sqrt{3} & 0 & -\sqrt{4} & .. \\ . & . & . & . & . & .. \end{pmatrix} \tag{4.72}$$

$$(H) = \hbar\omega \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \ldots. \\ 0 & \frac{3}{2} & 0 & 0 & \ldots. \\ 0 & 0 & \frac{5}{2} & 0 & \ldots. \\ 0 & 0 & 0 & \frac{7}{2} & \ldots. \\ . & . & . & . & \ldots. \end{pmatrix} \tag{4.73}$$

Alternate Method

A method due to Dirac avoids the dependence on the solutions of the differential equations of the Schrödinger picture for the evaluation of the matrix elements of operators in the Heisenberg picture. This method consists in finding suitable operators with which one can generate all the eigenvectors of the Hamiltonian from any given eigenvector. These eigenvectors will, then, define a representation. The Hamiltonian would, obviously, be diagonal in this representation.

Thus, the eigenvalue problem is automatically solved if we find the above operators. The method anticipates the technique of *field*, or *second, quantization* and is known also as the *method of second quantization* (see, Section 11.2).

The operators we seek are given by

$$\hat{a} = \frac{1}{\sqrt{2m\hbar\omega}}(m\omega\hat{x}_0 + i\hat{p}_0), \tag{4.74a}$$

$$\hat{a}^\dagger = \frac{1}{\sqrt{2m\hbar\omega}}(m\omega\hat{x}_0 - i\hat{p}_0). \tag{4.74b}$$

From Eqs. (4.74a, b), we get,

$$\hat{x}_0 = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a} + \hat{a}^\dagger), \tag{4.75a}$$

$$\hat{p}_0 = i\sqrt{\frac{m\hbar\omega}{2}}(\hat{a}^\dagger - \hat{a}). \tag{4.75b}$$

From the conditions,

$$[\hat{x}_0, \hat{p}_0] = i\hbar\hat{1};\ [\hat{x}_0, \hat{x}_0] = \hat{0} = [\hat{p}_0, \hat{p}_0],$$

we have

$$[\hat{a}, \hat{a}^\dagger] = \hat{1},$$

$$[\hat{a}, \hat{a}] = \hat{0} = [\hat{a}^\dagger, \hat{a}^\dagger]. \tag{4.76}$$

Substituting from Eqs. (4.75a, b) and (4.76) in eq. (4.57a), we get,

$$\hat{H} = \left(\hat{N} + \frac{\hat{1}}{2}\right)\hbar\omega, \tag{4.77}$$

where,

$$\hat{N} = \hat{a}^\dagger\hat{a}. \tag{4.78}$$

Thus, the problem of finding the eigenvalues and eigenvectors of $\hat{H}$ is reduced to the problem of determining the eigenvalues and eigenvectors of $\hat{N}$. Also, for the linear harmonic oscillator, the Hamiltonian, together with the parity operator, constitutes a complete set of operators. Therefore, $\hat{N}$ also will constitute such a complete set, so that determining the eigenvalues and eigenvectors of $\hat{N}$ solves the harmonic oscillator problem completely.

Now, from Eqs. (4.76) and (4.78), we have,

$$[\hat{N}, \hat{a}] = -\hat{a}, \tag{4.79a}$$

$$[\hat{N}, \hat{a}^\dagger] = +\hat{a}^\dagger, \tag{4.79b}$$

It is also easily shown that $\hat{N}$ is Hermitian.

Let ϕ_n represent a normalized eigenvector of $\hat{N}$ belonging to the eigenvalue n.

$$\hat{N}\phi_n = n\phi_n. \tag{4.80}$$

Then, from (4.79a), we have,

$$\hat{N}(\hat{a}\phi_n) = (n-1)(\hat{a}\phi_n), \tag{4.81}$$

which shows that $(\hat{a}\phi_n)$ is an eigenvector of $\hat{N}$ belonging to the eigenvalue $(n-1)$. Similarly, $(\hat{a}^r\phi_n)$ is an eigenvector of $\hat{N}$ belonging to the eigenvalue $(n-r)$. Now,

$$n \doteq (\phi_n, n\phi_n) = (\phi_n, \hat{N}\phi_n) = (\phi_n, \hat{a}^\dagger \hat{a}\phi_n)$$

$$= (\hat{a}\phi_n, \hat{a}\phi_n) = ||\hat{a}\phi_n||^2 \geq 0, \tag{4.82}$$

since the norm of a vector, in this case of $(\hat{a}\phi_n)$, cannot be negative. Therefore, the series,

$$\phi_n,\ \hat{a}\phi_n,\ \hat{a}^2\phi_n,\ \ldots,\ \hat{a}^r\phi_n \ldots \tag{4.83a}$$

should terminate as, otherwise, it would lead to a vector $\phi = (\hat{a})^s\phi_n$ for which the eigenvalue $(n - s)$ is negative. Let the last term of the series (4.83a) be denoted by ϕ_0. Then,

$$\hat{a}\phi_0 = 0. \tag{4.84}$$

Thus, the series (4.83a) correspond to the eigenvalues

$$n, (n-1), (n-2), \ldots, 0.$$

Similarly, from Eq. (4.79b), we get,

$$\hat{N}(\hat{a}^\dagger\phi_n) = (n+1)(\hat{a}^\dagger\phi_n),$$

and

$$\hat{N}(\hat{a}^\dagger)^r\phi_n = (n+r)(\hat{a}^\dagger)^r\phi_n,$$

showing that the series,

$$\hat{a}^\dagger\phi_n, (\hat{a}^\dagger)^2\phi_n, \ldots, (\hat{a}^\dagger)^r\phi_n, \ldots \tag{4.83b}$$

represents the eigenvalues

$$(n+1), (n+2), \ldots, (n+r), \ldots, +\infty.$$

Thus, the eigenvalue spectrum of $\hat{N}$ is given by the non-negative integers:

$$n = 0, 1, 2, \ldots, +\infty. \tag{4.85}$$

Any eigenvector of $\hat{N}$ can be reached from a given eigenvector ϕ by repeated application of either $\hat{a}$ or $\hat{a}^+$. Let us denote by the ket $|n>$ the normalized eigenvector of $\hat{N}$ belonging to eigenvalue n. That is

$$|n\rangle \equiv \phi_n, \tag{4.86a}$$

where

$$\langle n'|n\rangle = \delta_{n,n'}. \tag{4.87}$$

Then,

$$|n\rangle = C_n(\hat{a}^\dagger)^n|0\rangle = \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n|0\rangle, \tag{4.86b}$$

where

$$\hat{a}|0\rangle \equiv 0. \tag{4.82a}$$

Problem 4.7: Show that $C_n = \frac{1}{\sqrt{n!}}$ in Eq. (4.86b).

The set $\{|n\rangle\}$ of vectors, for n varying from 0 to $+\infty$, constitutes a complete, orthonormal set and, thus, defines a representation (called, the *occupation number representation*). The operator $\hat{N}$ is diagonal in this representation.

$$\langle n' | \hat{N} | n \rangle = n \delta_{n,n'} \tag{4.88}$$

Also,

$$\hat{a} | n \rangle = \frac{1}{\sqrt{n!}} \hat{a} (\hat{a}^\dagger)^n | 0 \rangle$$

$$= n \frac{1}{\sqrt{n!}} (\hat{a}^\dagger)^{n-1} | 0 \rangle = \sqrt{n} | n-1 \rangle, \tag{4.89a}$$

so that,

$$\langle n' | \hat{a} | n \rangle = \sqrt{n}\, \delta_{n',n-1}. \tag{4.90a}$$

Similarly,

$$\hat{a}^\dagger | n \rangle = \sqrt{n+1} | n+1 \rangle, \tag{4.89b}$$

and

$$\langle n' | \hat{a}^\dagger | n \rangle = \sqrt{n+1}\, \delta_{n',n+1} \tag{4.90b}$$

$$= \langle n | \hat{a} | n' \rangle.$$

Thus, the matrices (a), $(a^\dagger)$ and (N) representing the operators $\hat{a}, \hat{a}^\dagger$ and $\hat{N}$, respectively, are given by[7],

$$(a) = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & \ldots \\ 0 & 0 & \sqrt{2} & 0 & 0 & \ldots \\ 0 & 0 & 0 & \sqrt{3} & 0 & \ldots \\ 0 & 0 & 0 & 0 & \sqrt{4} & \ldots \\ . & . & . & . & & \ldots \end{pmatrix} \tag{4.91}$$

$$(a^\dagger) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \ldots \\ \sqrt{1} & 0 & 0 & 0 & 0 & \ldots \\ 0 & \sqrt{2} & 0 & 0 & 0 & \ldots \\ 0 & 0 & \sqrt{3} & 0 & 0 & \ldots \\ 0 & 0 & 0 & \sqrt{4} & 0 & \ldots \\ . & . & . & . & . & \ldots \end{pmatrix}, \tag{4.92}$$

7. Note that the first row corresponds to $n' = 0$ and the first column to $n = 0$.

$$(N) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 2 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 3 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 4 & \ldots \\ . & . & . & . & . & \ldots \end{pmatrix}. \tag{4.93}$$

From Eqs. (4.75a, b) and (4.77), we see that the matrices (4.91-93) are consistent with the matrices (4.71-73) representing $\hat{x}_0$, $\hat{p}_0$ and $\hat{H}$.

Problem 4.8: Show that $[\hat{H}, \hat{a}] = -(\hbar\omega)\hat{a}$ and $[\hat{H}, \hat{a}^\dagger] = (\hbar\omega)\hat{a}^\dagger$.

Problem 4.9: Show that the expectation value of the kinetic energy $\hat{T}$ in the state $|\, n >$ satisfies the Virial relationship,

$$\langle \hat{T} \rangle_n = \langle \hat{V} \rangle_n = \frac{1}{2} E_n.$$

Creation and Annihilation Operators

From Eqs. (4.77) and (4.88), we have, for the eigenvalue E_n of $\hat{H}$ in the state $|\, n\rangle$, the expression,

$$E_n = \left(n + \frac{1}{2}\right)\epsilon = n\,\epsilon + E_0, \tag{4.94}$$

where

$$\epsilon = \hbar\omega. \tag{4.95}$$

Thus,

$$(E_n - E_0) = n\,\epsilon, \quad n = 0, 1, 2, \ldots \tag{4.94a}$$

This equation permits the following interpretation: The oscillator in the state $|\, n\rangle$ is an assembly of n non-interacting particles, each of energy ϵ. The different states of the oscillator merely correspond to different numbers of the particles. From Eqs. (4.89a, b), we see that the operator $\hat{a}$ lowers the particle number by 1, whereas $\hat{a}^\dagger$ raises the particle number by 1. In other words, $\hat{a}^\dagger$ *creates* a particle while $\hat{a}$ *destroys*, or *annihilates*, one. $\hat{a}^\dagger$ and $\hat{a}$ are, therefore, called *creation* and *annihilation operators*, respectively. The relations, $\hat{a}\,|\,0\rangle = 0$, and $\hat{a}^\dagger\,|\,0\rangle = |\,1\rangle$, are consistent with this interpretation since a particle cannot be destroyed when no particle is present as in $|\,0\rangle$, but a particle can be created even when no particle is initially present. The interpretation of $\hat{N} \equiv \hat{a}^\dagger\hat{a}$, as the *number operator* is, then, suggested by Eq. (4.80).

It is in view of the foregoing interpretation that the representation defined by the basis vectors $|\, n\rangle$, $(n = 0, 1, 2, \ldots + \infty)$, is named the *occupation-number representation* (n is the number of particles occupying the state $|\, n\rangle$. It should, however, be remembered that the particle number 'n' is actually the quantum number characterising the state of excitation of the oscillator and it is only as a matter of convenience that it is called a particle-number.

4.2B The Hydrogen Atom

As another illustration of the application of the Schrödinger equation, we will take up the 3-dimensional problem of the hydrogen atom.

The hydrogen atom is a two-particle system, consisting as it does of a positively charged nucleus (the proton) and a negatively charged electron, moving under the influence of their mutual attraction. This means that ψ in Eq. (4.14b) is a function of $\mathbf{r}_1$ and $\mathbf{r}_2$, where $\mathbf{r}_1$ is the position vector of the electron and $\mathbf{r}_2$ that of the proton :

$$\psi = \psi(\mathbf{r}_1, \mathbf{r}_2, t). \tag{4.95}$$

Similarly,

$$\hat{H} \equiv \hat{H}(\mathbf{r}_1, \mathbf{r}_2, \nabla_1, \nabla_2, t)$$

$$= -\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 + \hat{V}(\mathbf{r}_1, \mathbf{r}_2, t), \tag{4.96}$$

where $$\nabla_i^2 \equiv \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$$

The Schrödinger equation for the hydrogen atom is thus,

$$i\hbar\frac{\partial}{\partial t}\psi(r_1, r_2, t) = \left[-\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 + \hat{V}(\mathbf{r}_1, \mathbf{r}_2, t)\right]\psi. \tag{4.97}$$

Now, in this case, $\hat{V}$ is derived from the Coulomb force, so that

$$\hat{V}(\mathbf{r}_1, \mathbf{r}_2, t) = -\frac{Ze^2}{|\mathbf{r}_1 - \mathbf{r}_2|} \equiv \hat{V}(|\mathbf{r}_1 - \mathbf{r}_2|), \tag{4.98}$$

where, the atomic number $Z = 1$ for the hydrogen atom. Writing $(\mathbf{r}_1 - \mathbf{r}_2) = \mathbf{r}$, we have,

$$\hat{V} \equiv \hat{V}(r) = -\frac{Ze^2}{r}, \tag{4.98a}$$

r being the separation between the particles. $\mathbf{r}$ is called the *relative co-ordinate.* Introducing also the *centre-of-mass co-ordinate* $\mathbf{R}$, defined by

$$M\mathbf{R} = m_1\mathbf{r}_1 + m_2\mathbf{r}_2, \tag{4.99}$$

where,

$$M = m_1 + m_2, \tag{4.100}$$

is the total mass of the system, we have,

$$\frac{1}{m_1}\nabla_1^2 + \frac{1}{m_2}\nabla_2^2 = \frac{1}{M}\nabla^2\mathbf{R} + \frac{1}{\mu}\nabla^2\mathbf{r} \tag{4.101}$$

where,

$$\mu = \frac{m_1 m_2}{m_1 + m_2}, \tag{4.102}$$

is called the *reduced mass* of the system.

Eq. (4.101) is easily derived by writing,

$$\nabla_1 \equiv \frac{\partial}{\partial \mathbf{r}_1} = \frac{\partial \mathbf{r}}{\partial \mathbf{r}_1}\frac{\partial}{\partial \mathbf{r}} + \frac{\partial \mathbf{R}}{\partial \mathbf{r}_1}\frac{\partial}{\partial \mathbf{R}}$$

$$= \frac{\partial}{\partial \mathbf{r}} + \frac{m_1}{M}\frac{\partial}{\partial \mathbf{R}} = \frac{m_1}{M}\nabla_{\mathbf{R}} + \nabla_{\mathbf{r}} \tag{4.103a}$$

and

$$\nabla_2 \equiv \frac{\partial}{\partial \mathbf{r}_2} = -\nabla_{\mathbf{r}} + \frac{m_2}{M}\nabla_{\mathbf{R}}. \tag{4.103b}$$

Substituting from Eqs. (4.98a) and (4.101) in Eq. (4.97), we have,

$$i\hbar\frac{\partial \Phi(\mathbf{r},\ \mathbf{R},\ t)}{\partial t} = \left[-\frac{\hbar^2}{2M}\nabla_{\mathbf{R}}^2 - \frac{\hbar^2}{2\mu}\nabla_{\mathbf{r}}^2 + V(r)\right]\Phi(\mathbf{r}, \mathbf{R}, t), \tag{4.97a}$$

where,

$$\Phi(\mathbf{r}, \mathbf{R}, t) = \psi(\mathbf{r}_1, \mathbf{r}_2, t). \tag{4.104}$$

Since the operator in the square bracket in (4.97a) is independent of time, we can write, according to Eq. (4.17b),

$$\Phi(\mathbf{r}, \mathbf{R}, t) = \phi(\mathbf{r}, \mathbf{R}) \exp\left[-(i/\hbar)\varepsilon t\right], \tag{4.104a}$$

where (see Eq. (4.18)),

$$\left[-\frac{\hbar^2}{2M}\nabla_{\mathbf{R}}^2 - \frac{\hbar^2}{2\mu}\nabla^2\mathbf{r} + V(r)\right]\phi(\mathbf{r},\ \mathbf{R}) = \varepsilon\phi(\mathbf{r},\ \mathbf{R}). \tag{4.97b}$$

Also, since each of the operators on L.H.S. of (4.97b) depends either on **r** or on **R**, but not on both **r** and **R**, a further separation of variables is possible in the form,

$$\phi(\mathbf{r},\ \mathbf{R}) = u(\mathbf{r})U(\mathbf{R}). \tag{4.104b}$$

Substituting this in Eq. (4.97b) and dividing throughout by $\phi(\mathbf{r},\ \mathbf{R})$ we get,

$$\left\{-\frac{\hbar^2}{2M}\frac{1}{U}\nabla_{\mathbf{R}}^2 U\right\} + \left[-\frac{\hbar^2}{2\mu}\frac{1}{u}\nabla_{\mathbf{r}}^2 u + V(r) - \varepsilon\right] = 0 \tag{4.105}$$

Since the expression in the curly bracket is purely a function of **R** and that in the square bracket a function of **r** only, each of the brackets should be equal to the same constant with opposite signs. Denoting this constant by E', we have,

$$-\frac{\hbar^2}{2M}\nabla_{\mathbf{R}}^2 U(\mathbf{R}) = E'U(\mathbf{R}), \tag{4.97c}$$

and

$$-\frac{\hbar^2}{2\mu}\nabla_{\mathbf{r}}^2 u(\mathbf{r}) + V(r)u(\mathbf{r}) = Eu(\mathbf{r}), \tag{4.97d}$$

where,

$$E = \varepsilon - E', \text{ or } \varepsilon = E + E'. \tag{4.106}$$

Equation (4.97c) represents the uniform motion of the centre of mass, E' being the kinetic energy associated with such a motion. Equation (4.97d), which represents the relative motion and, hence, depends on the internal structure of the atom, is the more interesting one. This equation has the appearance of the Schrödinger equation for a single particle of mass μ moving in a field represented by the potential $\hat{V}(r)$. We have, here, the well-known reduction of a two-body central force problem to a one-body problem.

The Relative Motion

Solution of the equation (4.97d) representing the motion of the electron[8] relative to the nucleus, would give us the energy levels and the wave functions of the atom. The method of solution to be described below, is applicable whenever the potential is spherically symmetric (that is, when V is a function of the radial co-ordinate r rather than of the vector $\mathbf{r}$).

In order to take full advantage of the spherical symmetry for the solution of the equation, we use spherical co-ordinates (r, θ, ϕ) instead of cartesian co-ordinates (x, y, z). The relationship between the two systems of co-ordinates is given by (see Fig. 5.1),

$$x = \mathbf{r} \sin\theta \cos\phi,$$

$$y = \mathbf{r} \sin\theta \sin\phi,$$

$$z = \mathbf{r} \cos\theta. \tag{4.107}$$

Then (see Eq. (5.41)),

$$\nabla_r^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right\}$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) - \frac{\hat{\mathbf{L}}^2}{\hbar^2 r^2}, \tag{4.108}$$

where, $\hat{\mathbf{L}}$ is the operator corresponding to the orbital angular momentum $\mathbf{r} \times \mathbf{p}$ (see Eq. 5.43)[9]. With the substitution of Eq. (4.108), Eq. (4.97d) takes the form,

$$\left[\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{2\mu r^2}{\hbar^2}\{E - V(r)\}\right] u(r,\theta,\phi) - (\hat{\mathbf{L}}^2/\hbar^2) u(r,\theta,\phi) = 0 \tag{4.109}$$

Since the terms in the square bracket are independent of the angular co-ordinates while $\mathbf{L}^2$ is independent of the radial co-ordinate, the solution would be of the form,

$$u(r,\theta,\phi) = R(r) Y(\theta,\phi). \tag{4.110}$$

Substituting in Eq. (4.109) and following the same procedure as in the case of (4.97b), we get the equations,

---

8. Since the mass of the proton is about 2000 times that of the electron, the reduced mass μ is only slightly smaller than the mass of the electron. Therefore, the motion in this case is nearly that of an electron around a fixed nucleus.

9. Eq. (4.108) corresponds to the resolution of the total linear momentum $\hat{p} \equiv -i\hbar\nabla_r$ of a particle in its orbital motion, into the radial component

$$\hat{\mathbf{p}}_r = -i\hbar \frac{\mathbf{r}}{\mathbf{r}^2}\frac{\partial}{\partial r} r = -i\hbar(\mathbf{r}/r)\left(\frac{\partial}{\partial r} + \frac{1}{r}\right)$$

and the perpendicular component $\hat{\mathbf{p}}_\perp = (\hat{\mathbf{L}}/\mathbf{r})$:

$$\mathbf{p} = \mathbf{p}_r + \mathbf{p}_\perp$$

$$\hat{L}^2 Y(\theta,\phi) = \hbar^2 \lambda Y(\theta,\phi), \tag{4.111}$$

and

$$\left[\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} + \frac{2\mu}{\hbar^2}(E - V(r)) - \frac{\lambda}{r^2}\right] R = 0, \tag{4.112}$$

where, λ is the separation constant.

From Eq. (4.111), we see that $Y(\theta,\phi)$ is the eigenvector of $\hat{L}^2$ and $\lambda\hbar^2$ the corresponding eigenvalue. That is (see Eqs (5.44), (5.46) and (5.56b)).

$$\lambda = l(l+1), \tag{4.111a}$$

$$Y(\theta,\phi) \equiv Y_{lm}(\theta,\phi) \tag{4.111b}$$

with

$$m = -l, -l+1, \ldots, +l, \tag{4.111c}$$

$$l = 0, 1, 2, \ldots + \infty. \tag{4.111d}$$

$Y_{lm}(\theta,\phi)$ is called the *spherical harmonic* of order l. Its properties are discussed in Section (5.4).

The Radial Equation

Substituting from (4.111a) and (4.98a) in (4.112), we have,

$$\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} + \left\{\frac{2\mu}{\hbar^2}\left(E + \frac{Ze^2}{r}\right) - \frac{l(l+1)}{r^2}\right\} R = 0$$

Or (multiplying throughout by r),

$$\left[r\frac{d^2}{dr^2} + 2\frac{d}{dr} + \left\{\frac{2\mu Ze^2}{\hbar^2} + \frac{2\mu E r}{\hbar^2} - \frac{l(l+1)}{r}\right\}\right] R = 0 \tag{4.112a}$$

Putting,

$$E = -\epsilon \; ; \alpha^2 = \frac{8\mu\epsilon}{\hbar^2} ; \rho = \alpha r, \tag{4.113}$$

We get,

$$\left[\rho\frac{d^2}{d\rho^2} + 2\frac{d}{d\rho} + \left\{\frac{2\mu Ze^2}{\alpha\hbar^2} - \frac{\rho}{4} - \frac{l(l+1)}{\rho}\right\}\right] y(\rho) = 0 \tag{4.112b}$$

where,

$$y(\rho) \propto R(r).$$

Equation (4.112b) is the differential equation for the *Associated Laguerre Function*, provided[10] (see Eq. (E.21)),

$$\frac{2\mu Ze^2}{\alpha\hbar^2} = n' - \frac{k-1}{2}, \tag{4.114a}$$

and

$$l(l+1) = \frac{k^2 - 1}{4}, \tag{4.114b}$$

10. If the conditions (4.114a, b) are not satisfied, Eq. (4.112b) has no solutions that are acceptable as wavefunctions of a physical system (that is, solutions that are finite, continuous and square integrable in the range $0 \le \rho \le \infty$).

where n' and k are both positive integers, and $n' \geq k$.

Equation (4.114b) can be written as

$$l(l+1) = \left(\frac{k-1}{2} + 1\right)\left(\frac{k-1}{2}\right),$$

so that,
$$\frac{k-1}{2} = l \geq 0, \tag{4.115a}$$

or
$$k = (2l+1), \text{ an odd integer.} \tag{4.115b}$$

Since $k \leq n'$, $\frac{k-1}{2} = l < n'$, so that

$$n = n' - \frac{k-1}{2} = (n' - l) \tag{4.116}$$

is a positive integer : $n = 1, 2, 3, \ldots$

From Eqs. (4.116), (4.114a) and (4.113), we have,

$$n^2 = \frac{\mu}{2} \cdot \left(\frac{Ze^2}{\hbar}\right)^2 \frac{1}{\epsilon}.$$

Thus, the energy eigenvalue E_n, corresponding to the quantum number n, is given by

$$E_n = -\epsilon_n = -\frac{\mu Z^2 e^4}{2\hbar^2 n^2} = -\frac{Z^2 e^2}{2a_0 n^2} \tag{4.117}$$

where,
$$a_0 = \frac{\hbar^2}{\mu e^2}. \tag{4.118}$$

a_0 has the dimension of length and is equal to the radius of the first orbit for hydrogen ($Z = 1$) in the Bohr atom model. It is called the *Bohr radius* of the atom.

The solution of Eq. (4.112b) is (see Eq. (E.20)),

$$y(\rho) = \mathcal{L}_{n',k}(\rho) = e^{-\rho/2}\rho^{(k-1)/2} L_{n'}^{k}(\rho)$$

$$= e^{-\rho/2}\rho^{l} L_{n+l}^{2l+1}(\rho), \tag{4.119}$$

where use has been made of Eqs. (4.115a, b) and (4.116). $L_n^{k'}(\rho)$ is the *Associated Laguerre Polynomial* of degree $(n' - k)$, given by Eq. (E.19):

$$L_{n+l}^{2l+1}(\rho) = \sum_{s=0}^{n-l-1} (-1)^{s+2l+1} \frac{[(n+l)!]^2 \rho^s}{(n-l-1-s)!(2l+1+s)!s!}. \tag{4.119a}$$

The radial wavefunction corresponding to the quantum numbers n and l (or n' and k), is given by

$$R_{nl}(r) = N_{nl} L_{n'k}(\rho) = N_{nl} \mathcal{L}_{(n+l),(2l+1)}(\rho), \tag{4.120a}$$

where, N_{nl} is a normalizing factor, such that

$$\int_0^\infty R_{nl}^2(r) r^2 dr = |N_{nl}|^2 \frac{1}{\alpha^3} \int_0^\infty \rho^2 |\mathcal{L}_{n',k}(\rho)|^2 d\rho = 1. \tag{4.121}$$

From Eq. (E.24a), we have,

$$\int_0^\infty \rho^2 \mid \mathcal{L}_{n',k}(\rho) \mid^2 d\rho = \frac{(n'!)^3(2n'-k+1)}{(n'-k)!}$$

$$= \frac{2n\{(n+l)!\}^3}{(n-l-1)!} \tag{4.121a}$$

Also, from Eqs. (4.113), (4.118) and (4.117), we get,

$$\alpha \equiv \alpha_n = \sqrt{\frac{-8E_n}{a_0(e^2)}} = \left(\frac{2Z}{na_0}\right) \tag{4.121b}$$

From Eqs. (4.121), (4.121a) and (4.121b), we have

$$N_{nl} = \pm\left[\left(\frac{2Z}{na_0}\right)^3 \frac{(n-l-1)!}{2n\{(n+l)!\}^3}\right]^{1/2}, \tag{4.122}$$

and $R_{nl}(r) = -\left[\left(\frac{2Z}{na_0}\right)^3 \cdot \frac{(n-l-1)!}{2n\{(n+l)!\}^3}\right]^{1/2} e^{-\rho/2}\rho^l L_{n+l}^{2l+1}(\rho)$; (4.120b)

where the negative sign is chosen so as to make R_{10} positive. The total wavefunction, Eq. (4.110), is given by

$$u_{nlm}(r,\theta,\phi) = R_{nl}(r)Y_{lm}(\theta,\phi), \tag{4.110a}$$

with $Y_{lm}(\theta,\phi)$ given by Eq. (5.56b). The quantum numbers n, l, m are, respectively, called the *total* quantum number, the *orbital* quantum number and the *magnetic* quantum number.

These quantum numbers determine the energy, the angular momentum and the angular momentum along the axis of quantization, respectively. It is seen from Eq. (4.119a) that, for a given n, the maximum value of l is $(n-1)$. Also, from the theory of angular momentum (see Eq. (5.49b)), we have that m varies from $-l$ to $+l$ for given l. Thus, the range of values of the three quantum numbers is as summarised below:

$$n = 1, 2, 3, \ldots, +\infty,$$

$$l = 0, 1, 2, \ldots, (n-1),$$

$$m = -l, -l+1, \ldots, +l. \tag{4.123}$$

It is customary to denote the l-value by an alphabet. The (spectroscopic) notation for different l-values are given below:

$$l = 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5$$

Notation: $s \quad p \quad d \quad f \quad g \quad h$ (4.124)

Using this notation and with the help of Table E.3, we list below some of the radial wavefunctions given by Eq. (4.120b):

$$R_{1s} = 2\left(\frac{1}{a}\right)^{3/2} e^{-r/a},$$

$$R_{2s} = \left(\frac{1}{2a}\right)^{3/2} [2-(r/a)]e^{-r/2a},$$

$$R_{2p} = \frac{1}{\sqrt{3}}\left(\frac{1}{2a}\right)^{3/2}(r/a)e^{-r/2a}, \tag{4.125}$$

$$R_{3s} = \frac{2}{27}\left(\frac{1}{3a}\right)^{3/2}[27 - .18(r/a) + 2(r/a)^2]e^{-r/3a},$$

$$R_{3p} = \frac{8}{27\sqrt{8}}\left(\frac{1}{3a}\right)^{3/2}(r/a)\,[6-(r/a)]e^{-r/3a},$$

$$R_{3d} = \frac{4}{27\sqrt{10}}\left(\frac{1}{3a}\right)^{3/2}\cdot(r/a)^2 e^{-r/3a},$$

where
$$a = (a_0 / Z). \tag{4.126}$$

The radial probability distribution is, however, proportional to $|\, rR_{nl}(r)\,|^2$. In Fig. 4.3, we have plotted $(r/a)R_{nl}(r)$ for some values of n and l.

We note from Eq. (4.120b), that $R_{nl}(r)$ has a node at $r = 0$ except for $l = 0$ (due to the factor ρ^l) and at r = infinity (due to the factor $e^{-\rho/2}$). Also, there are $(n - l - 1) = n_r$ nodes between $r = 0$ and $r = \infty$ (L_n^k being a polynomial of degree $n - k$). For this reason, n_r is called the *radial quantum number.*

Now, from Eqs. (4.110a) and (5.60c), we have,

$$\hat{P}u_{nlm}(r,\theta,\phi) = u_{nlm}(r,\pi-\theta,\pi+\phi)$$

$$= (-1)^l u_{nlm}(r,\theta,\phi), \tag{4.127}$$

where, $\hat{P}$ is the parity operator. Thus, the parity of the state is determined entirely by the orbital angular momentum. Since the energy depends only on the quantum number n (Eq. 4.117)), the *degree of degeneracy* of the level E_n is

$$\sum_{l=0}^{n-1}(2l+1) = n^2. \tag{4.128}$$

Of these, the $(2l+1)$-fold degeneracy (the m-degeneracy) associated with a given l, is a common feature of all central fields, arising from the inability of such fields to distinguish between different orientations in space. The l-degeneracy is, however, characteristic of the Coulomb field. In Fig. 4.4, we have plotted the energy levels along with the potential. The l-degeneracy gets removed in some hydrogen-like (alkali) atoms because of the screening of the coulomb field. The m-degeneracy can be removed by applying a non-central field such as a magnetic field in which each level (of a given l) splits up into $(2l+1)$ level (Zeeman Effect).

From Eqs. (4.110a), (4.120a), (4.122), (5.57), (E.25a) and definition (4.126), we get, for the expectation value of $\frac{1}{r}$, in the state $u_{nlm}(r,\theta,\phi)$, the expression,

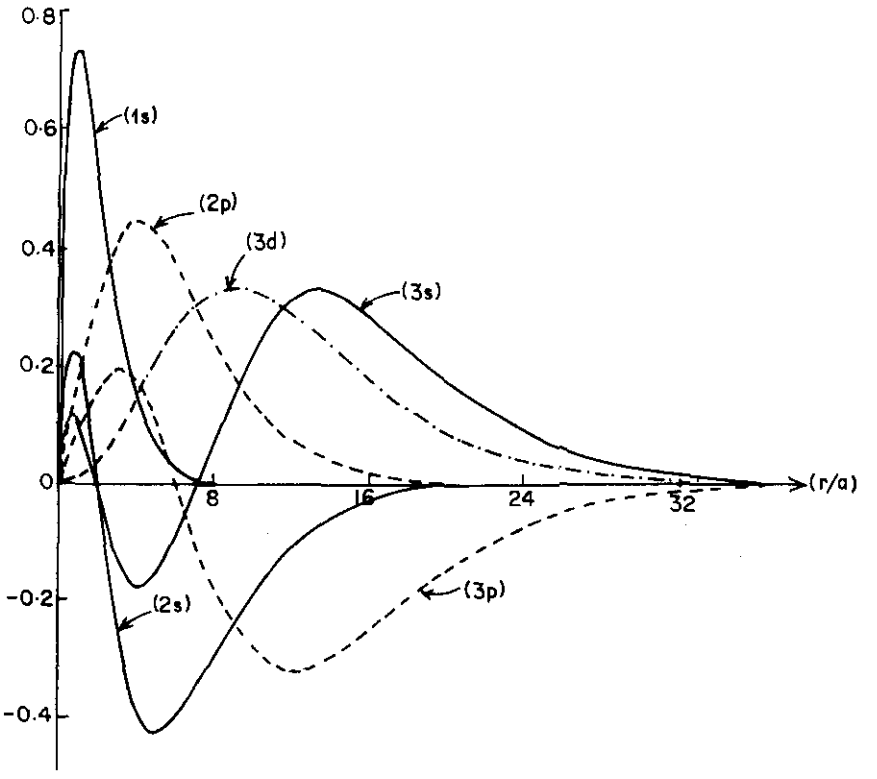


Fig. 4.3 The radial wave functions, $(r/a)R_{nl}(r)$, for the hydrogen atom.

$$\left\langle \frac{1}{r} \right\rangle_{nlm} \equiv \int \frac{1}{r} |u_{nlm}(r,\theta,\phi)|^2 r^2 dr d\Omega$$

$$= \int_0^\infty r R_{nl}^2(r) dr \int |Y_{lm}(\theta,\phi)|^2 d\Omega$$

$$= \frac{|N_{nl}|^2}{\alpha^2} \mathbf{I}_{n+l,n+1}^{(1)(2l+1)}$$

$$= \left(\frac{1}{n^2 a}\right). \tag{4.129a}$$

Similarly, using Eq. (E.25d), we get,

$$\langle (r/a) \rangle_{nlm} = \frac{1}{2}[3n^2 - l(l+1)]; \tag{4.129b}$$

and using (E.24),

$$\langle (r/a^2) \rangle_{nlm} = \frac{n^2}{2}[5n^2 - 3l(l+1) + 1]; \tag{4.129c}$$

and

$$\langle (r/a^s) \rangle_{nlm} = \left(\frac{n}{2}\right)^s \cdot \frac{(n-l-1)!}{2n[(n+l)!]^3} \mathbf{I}^{(s+2)(2l+1)}_{n+l,\,n+l}$$

$$= \left(\frac{n}{2}\right)^s \frac{(n-l-1)!}{2n(n+l)!} \sum_{r=0}^{s+1} \left\{ \frac{(s+1)!}{(s+1-r)!r!} \right\}^2 \frac{(n+l+s+1-r)!}{(n-l-1-r)!} \qquad (4.129d)$$

In the formula (4.129d), $s \geq -1$.

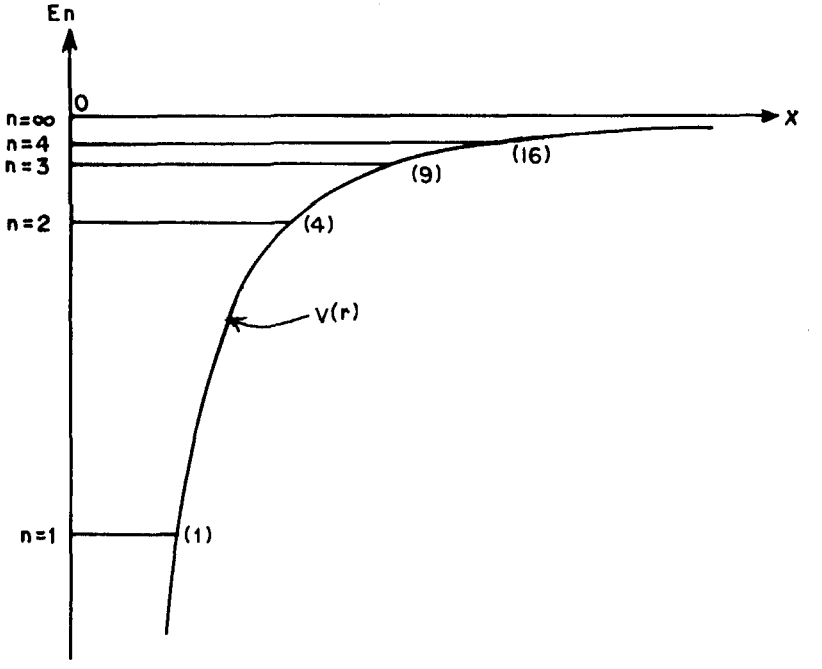


Fig. 4.4. Energy level diagram for the hydrogen atom shown in relation to the potential. The numbers in parenthesis on the right indicate the degeneracy of the level [Eqs. (4.117) and (4.128)].

Problem 4.10: Show from Eqs. (4.129a) and (4.117), that $\langle \hat{T} \rangle = -\frac{1}{2}\langle \hat{V} \rangle$, where $\hat{T}$ and $\hat{V}$, respectively, represent the kinetic and the potential energies of the atom.

REFERENCES

1. Roman, P. *Advanced Quantum Theory* (Addison-Wesley, Massachusetts 1962), Section 1.5.
2. Schiff, L.I. *Quantum Mechanics*, 3rd Edition (McGraw-Hill-Kogakusha, Tokyo 1968), Sections 13 and 16.

CHAPTER 5

THEORY OF ANGULAR MOMENTUM

Angular momentum plays a much more important role in quantum mechanics than in classical mechanics. This is, probably, due to the relatively greater importance of periodic motions in the former. Periodic motion can be envisaged as motion in a closed orbit which naturally involves angular momentum. The existence of the intrinsic angular momentum (spin) could be another reason. A third, and probably most important, reason is that angular momentum is quantized (unlike linear momentum).

5.1 THE DEFINITION

In classical mechanics, angular momentum of a particle about a point O (see Fig. 5.1) is defined as

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \tag{5.1}$$

where $\mathbf{r}$ is the position vector and $\mathbf{p}$ is the linear momentum of the particle. The corresponding quantum mechanical operator can be obtained from Eq. (5.1) by the application of Postulates IV and V: we replace the dynamical variables $\mathbf{r}$ and $\mathbf{p}$ by the corresponding operators. Then,

$$\begin{aligned} \hat{L}_x &= (\hat{y}\hat{p}_z - \hat{z}\hat{p}_y), \\ \hat{L}_y &= (\hat{z}\hat{p}_x - \hat{x}\hat{p}_z), \\ \hat{L}_z &= (\hat{x}\hat{p}_y - \hat{y}\hat{p}_x), \end{aligned} \tag{5.2a}$$

or, writing $x \equiv x_1$, $y \equiv x_2$, etc.

$$\hat{L}_i = \epsilon_{ijk}\, \hat{x}_j \hat{p}_k; \tag{5.2b}$$

where a repeated index is to be summed over.

Here,

$\epsilon_{ijk} = +1$, if ijk is an even permutation of the numbers 1, 2, 3.

$= -1$, if ijk is an odd permutation of 1, 2, 3.

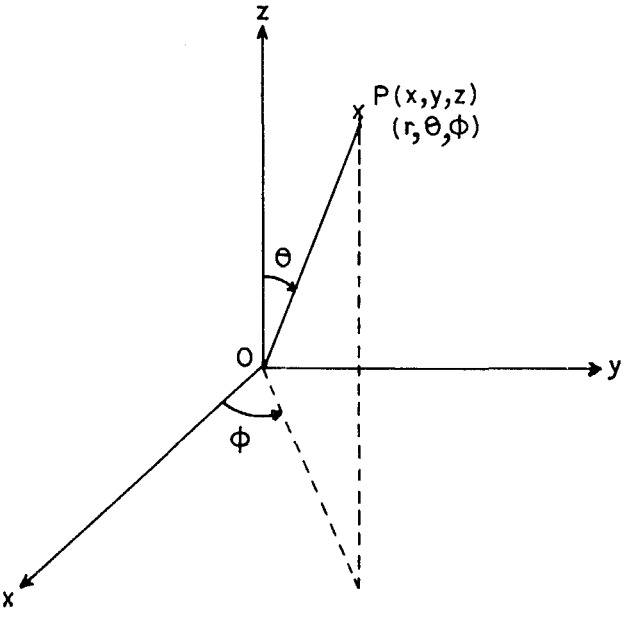


Fig. 5.1

$= 0$, if any two indices are equal.

Also, $$[\hat{x}_i, \hat{p}_j] = i\hbar\hat{\delta}_{ij},$$

$$[x_i, x_j] = [\hat{p}_i, \hat{p}_j] = \hat{0}. \tag{5.3}$$

From Eqs. (5.2b) and (5.3), with the help of the identity,

$$[ab, cd] \equiv a[b,c]d + [a,c]bd + c[a,d]b + ac[b,d],$$

we get,

$$[\hat{L}_i, \hat{L}_j] = i\hbar \in_{ijk} \hat{L}_k. \tag{5.4}$$

The square of $\hat{L}$ is defined by ,

$$\hat{\mathbf{L}}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2. \tag{5.5}$$

Then, $$[\hat{\mathbf{L}}^2, \hat{L}_k] = \hat{0}, \ (k = 1,2,3) \tag{5.6a}$$

or $$[\hat{\mathbf{L}}^2, \hat{\mathbf{L}}] = \hat{0}. \tag{5.6b}$$

Thus, the components of the angular momentum operator do not commute among themselves though they commute with the square of the angular momentum operator. As will be shown, the commutation relations, Eq. (5.4), determine the quantal properties of the angular momentum. That is, the eigenvalues and the eigenvectos of the angular momentum operator are completely determined by Eq. (5.4) and the general properties of the Hilbert space. Therefore, the commutation relations themselves are taken for the definition of the angular momentum operator in quantum mechanics.

Thus, *a vector operator $\hat{\mathbf{J}}$ is an angular momentum operator if its components are observables (hence, hermitian) and obey the commutation relations,*

$$[\hat{J}_x, \hat{J}_y] = i\hbar \hat{J}_z, \text{ cyclic} \tag{5.7a}$$

Since $$\hat{\mathbf{J}} \times \hat{\mathbf{J}} = [\hat{J}_y, \hat{J}_z]\mathbf{i} + [\hat{J}_z, \hat{J}_x]\mathbf{j} + [\hat{J}_x, \hat{J}_y]\mathbf{k},$$

the commutation relations can also be written as,

$$\hat{\mathbf{J}} \times \hat{\mathbf{J}} = i\hbar \hat{\mathbf{J}}. \tag{5.7b}$$

This definition enables one to treat entities which have no classical analogue, such as the spin and the iso-spin of elementary particles, on the same footing as angular momentum. The angular momentum represented by Eq. (5.2a) is called the *orbital* angular momentum.

Problem 5.1: Show that an operator which commutes with $\hat{J}_x$ and $\hat{J}_y$, commutes with $\hat{J}_z$ also.

Angular Momentum of a System of Particles

We postulate that the angular momentum operators referring to different particles commute:

$$[\hat{\mathbf{J}}_i, \hat{\mathbf{J}}_j] = \hat{0}, \quad i \neq j, \tag{5.8}$$

where the subscripts label the particles. Then, the operator corresponding to the total angular momentum of a system of N (non-interacting) particles, is given by,

$$\hat{\mathbf{J}} = \sum_{i=1}^{N} \hat{\mathbf{J}}_i. \tag{5.9}$$

It is easily verified that

$$[\hat{J}_x, \hat{J}_y] = i\hbar \hat{J}_z,$$

where $$\hat{J}_x = \sum_{i=1}^{N} \hat{J}_{ix}, \text{ etc.},$$

so that $\hat{\mathbf{j}}$ is, indeed, an angular momentum operator. Thus the vector-sum of a commuting set of angular momentum operators is an angular momentum operator.

5.2 EIGENVALUES AND EIGENVECTORS

Since the components of the angular momentum operator do not commute among themselves, we cannot find a common basis for all the three components. However, since $\hat{\mathbf{J}}^2$ commutes with $\hat{\mathbf{J}}$, we can have a common basis for $\hat{\mathbf{J}}^2$ and one of the components, say $\hat{J}_z$, of $\hat{\mathbf{J}}$.

Let $\{|\alpha jm\rangle\}$ represent such a common basis. Here, α represents the eigenvalues of operators (such as the Hamiltonian) which, together with $\hat{\mathbf{J}}^2$ and $\hat{J}_z$ form a complete set of commuting observables for the system. j labels the eigenvalues of $\hat{\mathbf{J}}^2$ and m those of $\hat{J}_z$. The vectors $|\alpha jm\rangle$ are orthonormal :

$$\langle \alpha' j' m' \mid \alpha j m \rangle = \delta_{\alpha\alpha'} \delta_{jj'} \delta_{mm'}. \tag{5.10}$$

For the sake of simplicity in writing, we will write $\mid jm >$ in place of $\mid \alpha jm \rangle$. By definition[1],

$$\hat{\mathbf{J}}^2 \mid jm \rangle = \lambda_j \hbar^2 \mid jm \rangle, \tag{5.11a}$$

$$\hat{J}_z \mid jm \rangle = m\hbar \mid jm \rangle. \tag{5.11b}$$

$\hat{\mathbf{J}}^2$, being the sum of squares of Hermitian operators, is positive definite. Therefore,

$$\lambda_j \equiv \frac{\langle jm \mid \hat{\mathbf{J}}^2 \mid jm \rangle}{\hbar^2} \geq 0. \tag{5.12}$$

Also,

$$\langle jm \mid \hat{\mathbf{J}}^2 \mid jm \rangle \equiv \langle \hat{\mathbf{J}}^2 \rangle = \langle \hat{J}_x^2 \rangle + \langle \hat{J}_y^2 \rangle + \langle \hat{J}_z^2 \rangle \geq \langle \hat{J}_z^2 \rangle,$$

so that, from (5.12),

$$\lambda_j \geq m^2 \geq 0. \tag{5.13}$$

It is convenient, at this stage, to introduce the non-Hermitian operators $\hat{J}_+$ and $\hat{J}_-$, defined by,

$$\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y. \tag{5.14}$$

In terms of $\hat{J}_+$ and $\hat{J}_-$,

$$\hat{J}_x = \frac{1}{2}(\hat{J}_+ + \hat{J}_-), \tag{5.15a}$$

$$\hat{J}_y = \frac{i}{2}(\hat{J}_- - \hat{J}_+), \tag{5.15b}$$

The following commutation relations for the set $\hat{J}_+, \hat{J}_-, \hat{J}_z$ and $\mathbf{J}^2$ are easily derived using the basic commutation relations (5.7a):

$$[\hat{J}_z, \hat{J}_\pm] = \pm \hbar \hat{J}_\pm, \tag{5.16a}$$

$$[\hat{J}_\pm, \hat{J}_\mp] = \pm 2\hbar \hat{J}_z, \tag{5.16b}$$

$$[\hat{\mathbf{J}}^2, \hat{J}_\pm] = [\hat{\mathbf{J}}^2, \hat{J}_z] = \hat{0}. \tag{5.16c}$$

Also,

$$\hat{\mathbf{J}}^2 = \frac{1}{2}(\hat{J}_+\hat{J}_- + \hat{J}_-\hat{J}_+) + \hat{J}_z^2, \tag{5.17a}$$

$$\hat{J}_-\hat{J}_+ = \hat{\mathbf{J}}^2 - \hat{J}_z(\hat{J}_z + \hbar), \tag{5.17b}$$

$$\hat{J}_+\hat{J}_- = \hat{\mathbf{J}}^2 - \hat{J}_z(\hat{J}_z - \hbar). \tag{5.17c}$$

---

1. We could write $\lambda_m \hbar$ instead of $m\hbar$ in Eq. (5.11b), but we have written $m\hbar$ in anticipation of the result.

Since $\hat{J}_z$ and $\hat{J}$ commute with $\mathbf{J}^2$, $\hat{J}_\pm \mid jm>$ are eigenvectors of $\mathbf{J}^2$ corresponding to the same eigenvalue as $\mid jm>$.

$$\mathbf{J}^2(\hat{J}_\pm \mid jm\rangle) = \lambda_j \hbar^2 (\hat{J}_\pm \mid jm\rangle). \tag{5.18a}$$

But, from Eqs. (5.16a) and (5.11b), we have,

$$\hat{J}_z(\hat{J}_\pm \mid jm\rangle) = (\hat{J}_\pm \hat{J}_z \pm \hbar \hat{J}_\pm) \mid jm\rangle$$

$$= (m \pm 1)\hbar(\hat{J}_\pm \mid jm\rangle). \tag{5.18b}$$

Thus, $\hat{J}_+ \mid jm\rangle$ is an eigenvector of $\hat{J}_z$ belonging to the eigenvalue $(m + 1)\hbar$ whereas $\hat{J}_- \mid jm\rangle$ is an eigenvector of $\hat{J}_z$ belonging to the eigenvalue $(m - 1)\hbar$:

$$\hat{J}_\pm \mid jm\rangle = \hbar c'_{\pm m} \mid jm \pm 1\rangle, \tag{5.19}$$

where $c'_{\pm m}$ is a scalar.

Since $\hat{J}_+$ raises the m-value of an eigenvector of $\mathbf{J}^2$ by 1 and $\hat{J}_-$ lowers the eigenvalue by 1, $\hat{J}_+$ and $\hat{J}_-$ are, respectively, called the *raising* and the *lowering* operators of angular momentum.

By repeated application of Eq. (5.19), we find,

$$(\hat{J}_\pm)^p \mid jm\rangle \propto \mid jm \pm p\rangle, \tag{5.19a}$$

where p is a positive integer or zero. This shows that we can reach any vector $\mid jm'\rangle$ by repeated application of $\hat{J}_+$ on $\mid jm\rangle$ if $m' - m$ = a positive integer, and by repeated application of $\hat{J}_-$ on $\mid jm\rangle$ if $m' - m$ = a negative integer. However, the series,

$$\hat{J}_+ \mid jm\rangle, \hat{J}_+^2 \mid jm\rangle, \ldots, \hat{J}_+^p \mid jm\rangle, \ldots \text{ and}$$

$$\hat{J}_- \mid jm\rangle, \hat{J}_-^2 \mid jm\rangle, \ldots, \hat{J}_-^p \mid jm\rangle, \ldots,$$

should terminate as, otherwise, we would have vectors $\mid jm'\rangle$ which violate the inequality (5.13), since λ_j is not changed by the application of $\hat{J}_\pm$ on $\mid jm\rangle$. Now, the series can terminate only if there is a value of m, say $m_>$, for which $\hat{J}_+ \mid jm_>\rangle = \mathbf{0}$, the null vector, and another value $m_<$ for which $\hat{J}_- \mid jm_<\rangle = \mathbf{0}$. Since $\mid jm_>\rangle$ and $\mid jm_<\rangle$ are obtained from $\mid jm\rangle$ by repeated application of $\hat{J}_+$ and $\hat{J}_-$ respectively, we have,

$$m_> - m_< = \text{ a positive integer or zero.} \tag{5.20}$$

Now the vector $\hat{J}_+ \mid jm_>\rangle = \mathbf{0}$, when its norm is zero. That is, when,

$$(\hat{J}_+ \mid jm_>\rangle)^* (\hat{\mathbf{J}}_+ \mid jm_>\rangle) = 0,$$

or

$$\langle jm_> \mid \hat{\mathbf{J}}_- \hat{\mathbf{J}}_+ \mid jm_>\rangle = 0, \tag{5.21}$$

since $\hat{J}_+^\dagger = \hat{J}_-$.

Using the expression (5.17b) for $\hat{J}_-\hat{J}_+$ and from Eqs. (5.10), (5.11a, b) and (5.21), we get,

$$\lambda_j = m_>(m_> + 1). \tag{5.22a}$$

Similarly, $\hat{J}_- | jm_< > = \mathbf{0},$

if $$\lambda_j = m_<(m_< - 1) \tag{5.22b}$$

Combining Eqs. (5.22a) and (5.22b), we have,

$$m_>(m_> + 1) = m_<(m_< - 1),$$

or, $$(m_> + m_<)(m_> - m_< + 1) = 0.$$

Thus, either $$m_> = m_< - 1, \tag{5.23a}$$

or, $$m_> = -m_<. \tag{5.23b}$$

Condition (5.23a) is ruled out by Eq. (5.20). Hence, only (5.23b) is acceptable.

Now, λ_j, being the eigenvalue of $\hat{\mathbf{J}}^2$, depends only on j (by our definition of $| jm \rangle$, so that, according to Eqs. (5.22a, b), $m_>$ and $m_<$ should be functions of j only. The choice,

$$m_> = j, \tag{5.24}$$

meets with these conditions. Then,

$$m_< = -m_> = -j.$$

So that, from Eq. (5.20),

$$m_> - m_< = 2j = \text{ a positive integer or zero.}$$

Thus,

$$j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots, +\infty. \tag{5.25}$$

The eigenvalue λ_j, according to Eq. (5.22a), is given by

$$\lambda_j = j(j + 1). \tag{5.26}$$

m can have any value between j and $-j$ such that

$j - m =$ a positive integer or zero. That is,

$$m = -j, -j + 1, \ldots, +j \tag{5.27}$$

Since the value of j, the maximum value of the projection of the angular momentum vector on the z-axis, fixes the length of the angular momentum vector uniquely, the latter is usually specified by its j value. Thus the statement: "the angular momentum of the particle is 3/2", means that the angular momentum vector is of length $\sqrt{\frac{3}{2}\cdot\frac{5}{2}}\hbar$. This length can, however, never be observed directly.

For a given value of j, there are $(2j + 1)$ linearly independent vectors $| jm >$, corresponding to the $(2j + 1)$ different values of m given by Eq. (5.27), which are common eigenvectors of $\hat{\mathbf{J}}^2$ and $\hat{J}_z$. If one of these vectors is given, the others can

be generated from it by the repeated application of $\hat{J}_+$ and $\hat{J}_-$. Let $|jm>$ represent the given vector, where $-j < m < j$. Then, from Eqs. (5.19), (5.17b), (5.11a, b) and (5.26), we get,

$$||\hat{J}_+|jm>|| = \hbar_1|c^j_{+m}| = \sqrt{j(j+1)-m(m+1)}\hbar$$

$$= \sqrt{(j-m)(j+m+1)}\hbar$$

so that,
$$c^j_{+m} = e^{i\delta}\sqrt{(j-m)(j+m+1)} \tag{5.28a}$$

Similarly,
$$c^j_{-m} = e^{i\gamma}\sqrt{(j+m)(j-m+1)}. \tag{5.28b}$$

Here, δ and γ are real scalars, independent of j and m. From Eqs. (5.19) and (5.28a, b), we have

$$\hat{J}_+|jm> = e^{i\delta}\{(j-m)(j+m+1)\}^{1/2}\hbar|jm+1> \tag{5.29a}$$

and
$$\hat{J}_-|jm> = e^{i\gamma}\{(j+m)(j-m+1)\}^{1/2}\hbar|jm-1>. \tag{5.29b}$$

Multiplying Eq. (5.29a) with $\hat{J}_-$ and using Eq. (5.29b),

$$\hat{J}\,\hat{J}_-|jm> = e^{i(\delta+\gamma)}\{(j-m)(j+m+1)\}\hbar^2|jm>.$$

But, using the expression (5.17b) for $\hat{J}_-\hat{J}_+$, we get,

$$\hat{J}_-\hat{J}_+|jm> = \{(j-m)(j+m+1)\}\hbar|jm>.$$

Hence,
$$e^{i(\delta+\gamma)} = 1,$$

or,
$$\gamma = -\delta. \tag{5.30}$$

Thus, the choice of δ fixes the phase of all the vectors relative to $|jm>$. The phase of $|jm>$ itself is, however, arbitrary. Following the usual practice, we put $\delta = 0$. Then,

$$\hat{J}_+|jm> = \{(j-m)(j+m+1)\}^{1/2}\hbar|jm+1>, \tag{5.31a}$$

$$\hat{J}_-|jm> = \{(j+m)(j-m+1)\}^{1/2}\hbar|jm-1>, \tag{5.31b}$$

Problem 5.2: Deduce the following relationships:

(a)
$$|j\pm m> = \frac{1}{\hbar^{j-m}}\left\{\frac{(j+m)!}{(2j)!(j-m)!}\right\}^{1/2}(\hat{J}_\pm)^{j-m}|j\pm j>.$$

(b)
$$|j\pm j> = \frac{1}{(\hbar)^{j-m}}\left\{\frac{(j+m)!}{(2j)!(j-m)!}\right\}^{1/2}(\hat{J}_\pm)^{j-m}|j\pm m>.$$

5.3 MATRIX REPRESENTATION

The vectors $|jm>$, for $m = -j$ to j, constitute a basis for a $(2j+1)$ – dimensional subspace of the Hilbert space of the system. The components of $\hat{\mathbf{J}}$ as well as $\hat{\mathbf{J}}^2$ are represented by Hermitian matrices in this space, where the rows and columns are labelled by the $(2j+1)$ values of m. Thus,

$$J^2_{m'm} \equiv < jm' \mid \hat{\mathbf{J}}^2 \mid jm > = j(j+1)\hbar^2\delta_{m',m} \tag{5.32a}$$

$$(J_z)_{m',m} \equiv < jm' \mid \hat{J}_z \mid jm > = m\hbar\delta_{m'm}. \tag{5.32b}$$

From Eqs. (5.31 a, b), we have,

$$(J_+)_{m',m} = \sqrt{(j-m)(j+m+1)}\,\hbar\delta_{m',m+1}, \tag{5.32c}$$

$$(J_-)_{m'm} = \sqrt{(j+m)(j-m+1)}\,\hbar\delta_{m',m-1}, \tag{5.32d}$$

The matrices representing $\hat{J}_x$ and $\hat{J}_y$ could be obtained from Eqs. (5.15 a, b) and (5.32 c, d):

$$J_x = \frac{1}{2}(J_+ + J_-), \tag{5.32e}$$

$$J_y = \frac{i}{2}(J_- - J_+). \tag{5.32f}$$

As expected, J^2 and J_z are diagonal in this representation.

Problem 5.3: If J_x and J_z are real matrices show that J_y is a purely imaginary matrix.

Problem 5.4: Show that $Tr(J_\mu) = 0$, $(\mu = x, y, z)$.

Problem 5.5: Obtain the angular momentum matrices corresponding to $j = 1$.

Pauli Spin Matrices

When $j = \frac{1}{2}$, we have, from Eqs. (5.32 b-d),

$$(J_z)_{1/2,1/2} = -(J_z)_{-1/2,-1/2} = \frac{\hbar}{2},$$

$$(J_z)_{-1/2,1/2} = (J_z)_{1/2,-1/2} = 0,$$

$$(J_\pm)_{1/2,1/2} = (J_\pm)_{-1/2,-1/2} = 0,$$

$$(J_+)_{-1/2,1/2} = (J_-)_{1/2,-1/2} = 0,$$

$$(J_+)_{1/2,-1/2} = (J_-)_{-1/2,1/2} = \hbar.$$

Hence, the matrices J_z, J_+ and J_- are given by

$$J_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix};$$

$$J_+ = \hbar\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \; J_- = \hbar\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Then, from Eqs. (5.32 e, f),

$$J_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};\ J_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

The Pauli spin matrices[2] $\sigma_x, \sigma_y, \sigma_z$ are defined by

$$J_\mu = \frac{\hbar}{2}\sigma_\mu,\ (\mu = x, y, z). \tag{5.33a}$$

Thus,
$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};\ \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix};\ \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{5.34}$$

We will see, in the next section, that the angular momentum arising from the orbital motion of a particle corresponds to integral values of j. Therefore, $j = \frac{1}{2}$ corresponds to the intrinsic angular momentum, or *spin*, of a particle.[3] Denoting the spin vector by s, and the vector whose components are $\sigma_x, \sigma_y, \sigma_z$ by σ, we can write Eq. (5.33a) as,

$$\mathbf{s} = \frac{\hbar}{2}\boldsymbol{\sigma}. \tag{5.33b}$$

The following properties of the Pauli spin matrices are easily verified (For ease of writing, we replace x, y, z by 1, 2, 3, respectively).

$$[\sigma_i, \sigma_j] = 2i\varepsilon_{ijk}\sigma_k, \tag{5.35a}$$

$$\{\sigma_i, \sigma_j\} = 2\delta_{ij}, \tag{5.35b}$$

$$\sigma_i\sigma_j = \delta_{ij} + i\varepsilon_{ijk}\sigma_k, \tag{5.35c}$$

$$\sigma_1\sigma_2\sigma_3 = i, \tag{5.35d}$$

$$\mathrm{Tr}(\sigma_k) = 0, \tag{5.35e}$$

$$\det(\sigma_k) = -1. \tag{5.35f}$$

Here $\{a, b\} \equiv ab + ba$, is the *anticommutator* of a and b.

Problem 5.6: If $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are vector operators such that

$$[\hat{\sigma}, \hat{\mathbf{A}}] = [\hat{\sigma}, \hat{\mathbf{B}}] = \hat{0}, \text{show that,}$$

$$(\hat{\sigma}\cdot\hat{\mathbf{A}})(\hat{\sigma}\cdot\hat{\mathbf{B}}) = (\hat{\mathbf{A}}\cdot\hat{\mathbf{B}}) + i\hat{\sigma}\cdot(\hat{\mathbf{A}}\times\hat{\mathbf{B}})$$

Problem 5.7: Write down the Pauli spin matrices in a representation in which σ_y is diagonal.

---

2. Pauli, Wolfgang [Z. f. Physik, **43**, 601 (1927)] was the first to introduce them.
3. The hypothesis of an intrinsic angular momentum for the electron was put forward by G.E. Uhlenbeck and S. Goudsmit [*Die Naturwissenschaften* **13**, 953 (1925)]. It is now recognised that spin, like electric charge, is one of the intrinsic attributes of all elementary particles. The spin could be half integral or integral, including zero.

Problem 5.8: Show that if A is a matrix such that

$$\{A, \sigma_\mu\} = 0, \ \mu = x, y, z,$$

then, A is null.

The Spin-Eigenvector

The eigenvectors corresponding to spin are given by[4]

$$|\alpha\rangle \equiv |\frac{1}{2}\frac{1}{2}\rangle, \ \left(j = \frac{1}{2}, m = +\frac{1}{2}\right), \tag{5.36a}$$

and

$$|\beta\rangle \equiv |\frac{1}{2}-\frac{1}{2}\rangle, \ \left(j = \frac{1}{2}, m = -\frac{1}{2}\right). \tag{5.36b}$$

$|\alpha\rangle$ represents the *spin-up* state and $|\beta\rangle$ the *spin-down* state:

$$\hat{\sigma}_z |\alpha\rangle = +|\alpha\rangle, \tag{5.37a}$$

$$\hat{\sigma}_z |\beta\rangle = -|\beta\rangle, \tag{5.37b}$$

In matrix notation the eigenvectors are

$$\alpha \equiv \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \text{ and } \beta \equiv \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

where

$$\sigma_z \alpha = \alpha; \ \sigma_z \beta = -\beta.$$

Using expression (5.34) for σ_z, we get

$$\alpha \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \ \beta \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{5.36c}$$

α and β represent, what are called, *pure states*. In an ensemble of spin $\frac{1}{2}$ particles, it is unlikely that all the particles are spin-up or all spin-down. It is more likely that some of the particles are spin-up and the others are spin-down. The spin-wavefunction, or *spinor*, χ corresponding to an ensemble is, therefore, a linear superposition of α and β:

$$\chi = c_1\alpha + c_2\beta = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \tag{5.38}$$

$$= \chi_+ + \chi_-,$$

where

$$|c_1|^2 + |c_2|^2 = 1 \tag{5.39}$$

5.4 ORBITAL ANGULAR MOMENTUM

The results in the previous two sections have been deduced from the general definition, (5.7b), of angular momentum and the general properties of vectors and

4. α, here, is not to be confused with the α in Eq. (5.10).

operators in the Hilbert space. We will now discuss the properties of the angular momentum represented by Eq. (5.2a), which, as its classical counterpart, Eq. (5.1), indicates, arises from the orbital motion of particles.

In the co-ordinate representation, we have [see Eq. (3.18[1])], $\hat{\mathbf{p}} \equiv -i\hbar\nabla$, so that from Eq (5.2a), we have[5],

$$\begin{aligned}
\hat{L}_x &= -i\hbar\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right), \\
\hat{L}_y &= -i\hbar\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z}\right), \\
\hat{L}_z &= -i\hbar\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right).
\end{aligned} \tag{5.40}$$

Angular momentum, as we will see in Section 5.6, is intimately related to rotations of a physical system in space. It would, therefore, be advantageous to use spherical co-ordinates (r, θ, ϕ) in place of cartesian co-ordinates (x, y, z) in Eq. (5.40). Using the relationships (4.107) and

$$\begin{aligned}
\frac{\partial r}{\partial x} &= \frac{x}{r}, \ \frac{\partial \theta}{\partial x} = \frac{x\cot\theta}{r^2}, \ \frac{\partial \phi}{\partial x} = -\frac{\sin\theta}{r\sin\theta}, \\
\frac{\partial r}{\partial y} &= \frac{y}{r}, \ \frac{\partial \theta}{\partial y} = \frac{y\cot\theta}{r^2}, \ \frac{\partial \phi}{\partial y} = -\frac{\cos\theta}{r\sin\theta}, \\
\frac{\partial r}{\partial z} &= \frac{z}{r}, \ \frac{\partial \theta}{\partial z} = -\frac{1}{r}\sin\theta, \ \frac{\partial \phi}{\partial z} = 0,
\end{aligned}$$

we have,

$$\begin{aligned}
\frac{\partial}{\partial x} &= \frac{x}{r}, \left[\frac{\partial}{\partial r} + \frac{1}{r}\cot\theta, \ \frac{\partial}{\partial \theta} - \frac{\sin\phi}{x\sin\theta}\frac{\partial}{\partial \phi}\right], \\
\frac{\partial}{\partial y} &= \frac{y}{r}, \left[\frac{\partial}{\partial r} + \frac{1}{r}\cot\theta, \ \frac{\partial}{\partial \theta} + \frac{\cos\phi}{x\sin\theta}\frac{\partial}{\partial \phi}\right], \\
\frac{\partial}{\partial z} &= \frac{z}{r}\left[\frac{\partial}{\partial r} - \frac{\sin\theta}{z}\frac{\partial}{\partial \theta}\right]
\end{aligned} \tag{5.41}$$

so that,

$$\hat{L}_x = i\hbar\left[\sin\phi\frac{\partial}{\partial\theta} + \cot\theta\cos\phi\frac{\partial}{\partial\phi}\right], \tag{5.42a}$$

---

5. With the differential operator form for $\hat{\mathbf{p}}$, the commutation rules $[\hat{x}, \hat{p}_x] = i\hbar\hat{1}$, etc., are automatically implied without making a distinction between $\hat{x}$ and x. Therefore, in this section, we will write x, y, z in place of $\hat{x}, \hat{y}, \hat{z}$.

$$\hat{L}_y = i\hbar\left[-\cos\phi\frac{\partial}{\partial\theta} + \cot\theta\sin\phi\frac{\partial}{\partial\phi}\right], \tag{5.42b}$$

$$\hat{L}_z = -i\hbar\frac{\partial}{\partial\phi}, \tag{5.42c}$$

and

$$\hat{\mathbf{L}}^2 = \hat{L}_x^2 + L_y^2 + L_z^2,$$

$$= -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]. \tag{5.43}$$

Problem 5.9: Establish the following commutation relations for the components of the angular and linear momenta:

$$[\hat{L}_i, \hat{p}_j] = i\hbar \in_{ijk} \hat{p}_k.$$

Hence show that $[\hat{\mathbf{L}}, \hat{\mathbf{p}}^2] = \hat{0}$

We will denote the eigenvectors of $\hat{\mathbf{L}}^2$ and $\hat{L}_z$ by $|\, lm >$, so that,

$$\hat{\mathbf{L}}^2 |\, lm > = l(l+1)\hbar^2 |\, lm >, \tag{5.44}$$

$$\hat{L}_z |\, lm > = m\hbar |\, lm > . \tag{5.45}$$

Now $\hat{\mathbf{L}}^2$ and $\hat{L}_z$ are purely functions of θ and ϕ. Therefore, in the co-ordinate representation, the eigenvectors $|\, lm >$ also should be a function of θ and ϕ only:

$$< \mathbf{r} |\, lm > \equiv Y_{lm}(\theta, \phi) \tag{5.46}$$

From Eqs. (5.42c), (5.45) and (5.46), we have,

$$-i\hbar\frac{\partial}{\partial\phi}Y_{lm}(\theta,\phi) = m\hbar Y_{lm}(\theta,\phi).$$

Integrating with respect to ϕ, we get,

$$Y_{lm}(\theta,\phi) = f_{lm}(\theta)e^{im\phi}, \tag{5.47}$$

where f_{lm} is independent of ϕ.

Now, the wavefunction Y_{lm} should be a single-valued function of θ and ϕ, so that,

$$Y_{lm}(\theta, \phi + 2\pi) = Y_{lm}(\theta, \phi). \tag{5.48}$$

Eqs. (5.47) and (5.48) yield,

$$e^{i2\pi m} = 1,$$

or, $m = \pm n$, where n is a positive integer or zero.

Hence, $l \equiv$ largest value of m

$=$ a positive integer or zero.

That is,

$$l = 0, 1, 2, \ldots, +\infty \tag{5.49a}$$

$$m = -l, -l+1, \ldots, +l. \tag{5.49b}$$

Thus, in the case of orbital angular momentum only the integer values are allowed in Eq. (5.25).

In order to determine $f_{lm}(\theta)$ in Eq. (5.47), we have to solve the equation,

$$\hat{L}^2 Y_{lm}(\theta, \phi) - l(l+1)\hbar^2 Y_{lm}(\theta, \phi) = 0.$$

With the help of Eqs. (5.44) and (5.47), and with the substitution, $\xi = \cos\theta$, this equation could be reduced to,

$$\left[(1-\xi^2)\frac{d^2}{d\xi^2} - 2\xi\frac{d}{d\xi} + \left\{l(l+1) - \frac{m^2}{1-\xi^2}\right\}\right] w_{lm}(\xi) = 0, \tag{5.50}$$

The two independent solutions of Eq. (5.50) are [see Eq. (E.35)],

$$w_{lm}^{(1)}(\xi) = P_l^m(\xi) = (1-\xi^2)^{m/2}\frac{d^m P_l(\xi)}{d\xi^m}, \tag{5.51a}$$

and

$$w_{lm}^{(2)}(\xi) = Q_l^m(\xi) = (1-\xi^2)^{m/2}\frac{d^m Q_l(\xi)}{d\xi^m}, \tag{5.51b}$$

where $P_l(\xi)$ and $Q_l(\xi)$ are solutions of the Legendre's differential equation,

$$\left[(1-\xi^2)\frac{d^2}{d\xi^2} - 2\xi\frac{d}{d\xi} + l(l+1)\right] w_l(\xi) = 0, \tag{5.52}$$

and are known as *Legendre polynomials* of the first and the second kind, respectively, of degree l. $P_l^m(\xi)$ and $Q_l^m(\xi)$ are, respectively, the *Associated Legendre functions* of the first and the second kind. $Q_l^m(\xi)$ is not acceptable as a wavefunction, since it is not finite at all points in the interval $-1 \le \xi \le 1$. Thus, the solution that is related to $f_{lm}(\theta)$ is given by Eq. (5.51a). Using the Rodrigue's formula [Eq. (E.26b)],

$$P_l(\xi) = \frac{1}{2^l l!}\frac{d^l}{d\xi^l}(\xi^2 - 1)^l, \tag{5.53}$$

for the Legendre polynomial, the solution can be written as,

$$P_l^m(\xi) = (1-\xi^2)^{m/2}\left(\frac{1}{2^l l!}\frac{d^{l+m}}{d\xi^{l+m}}(\xi^2-1)^l\right) \tag{5.54}$$

The properties of P_l^m are listed in Section (E.3).

The eigenvector $< \mathbf{r} \mid lm >$ is, thus, given by,

$$< \mathbf{r} \mid lm > \equiv Y_{lm}(\theta, \phi) = c_{lm} P_l^m(\cos\theta) e^{im\phi}, \tag{5.55c}$$

where, c_{lm} is a normalising constant, to be determined from the condition,

$$< lm \mid lm > \equiv \int Y^*_{lm}(\theta,\phi) Y_{lm}(\theta,\phi) d\Omega = 1, \tag{5.66}$$

where $d\Omega = \sin\theta d\theta d\phi = -d(\cos\theta)d\phi$, is an element of solid angle. Now, from Eq. (5.55a),

$$\int Y^*_{lm}(\theta,\phi) Y_{lm}(\theta,\phi) d\Omega = |c_{lm}|^2 \int_{-1}^{+1} \{P_1^m(\cos\theta)\}^2 d(\cos\theta)$$

$$\times \int_0^{2\pi} e^{-im\phi} e^{im\phi} d\phi$$

$$= |c_{lm}|^2 \cdot \frac{4\pi}{2l+1} \cdot \frac{(l+m)!}{(l-m)!}, \qquad \text{using Eq. (4.40).}$$

Thus, $$|c_{lm}| = \left[\frac{2l+1}{4\pi} \cdot \frac{(l-m)!}{(l+m)!}\right]^{1/2},$$

or $$c_{lm} = e^{i\delta}\left\{\frac{2l+1}{4\pi} \cdot \frac{(l-m)!}{(l+m)!}\right\}^{1/2} \tag{5.57}$$

The factor $e^{i\delta}$ represents the arbitrary phase factor we mentioned in Section (5.2). We choose,

$$e^{i\delta} = (-1)^m. \tag{5.58}$$

Then, $$Y_{lm}(\theta,\phi) = (-1)^m \left[\frac{2l+1}{4\pi} \cdot \frac{(l-m)!}{(l+m)!}\right]^{1/2} P_l^m(\cos\theta) e^{im\phi}$$

$$= (-1)^{l+m} \frac{1}{2^l l!} \left\{\frac{(2l+1)!}{4\pi} \frac{(l-m)!}{(l+m)!}\right\}^{1/2} \sin^{|m|}\theta$$

$$\times \left(\frac{d}{d(\cos\theta)}\right)^{l+m} \sin^{2l}\theta . e^{im\phi}. \tag{5.55b}$$

$Y_{lm}(\theta,\phi)$ is called the *spherical*[6] *harmonic* of order l.

We list below some of the important properties of the spherical harmonics; which follow from Eqs. (5.55b) and (E.37), (E.38) and (E.40).

$$Y^*_{lm}(\theta,\phi) = (-1)^m Y_{l-m}(\theta,\phi). \tag{5.59a}$$

$$= (-1)^m Y_{lm}(-\theta,-\phi). \tag{5.59b}$$

---

6. The Laplace operator ∇^2, in spherical co-ordinates, is given by

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) - \frac{\hat{\mathbf{L}}^2}{\hbar^2 r^2}$$

Thus, $r^l Y_{lm}(\theta,\phi)$ are the solutions of the Laplace equation $\nabla^2\Phi = 0$. The Y_{lm}'s are, therefore, the solutions of the Laplace equation on the unit sphere ($r = 1$). Hence the name spherical harmonics.

Orthonormality:

$$\int Y^*_{l'm'}(\theta,\phi)Y_{lm}(\theta,\phi)d\Omega = \delta_{ll'}\delta_{mm'} \tag{5.60}$$

Parity: If $\hat{P}$ represents the parity operator,

$$\hat{P}Y_{lm}(\theta,\phi) = (-1)^l Y_{lm}(\theta,\phi). \tag{5.61}$$

Thus, the parity of Y_{lm} is $(-1)^l$; which is even for even values of l and odd for odd values of l. This result is easily derived as follows:

The parity operation is the reflection of the co-ordinate axes at the origin. Therefore under this operation (see Fig. 5.2)

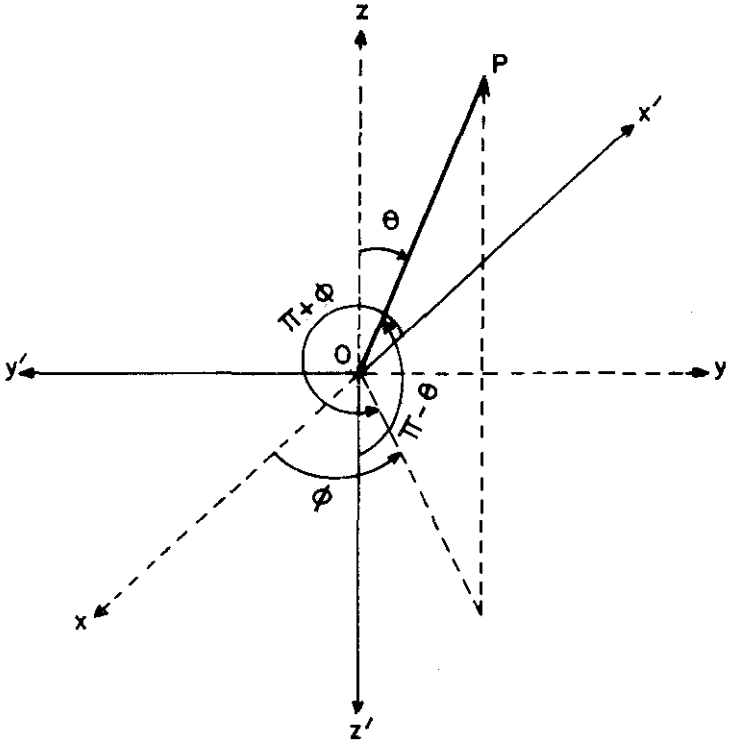


Fig. 5.2. The effect of parity operation on the angular co-ordinates (θ,ϕ).

$$\theta \to \pi - \theta, \quad \phi \to \pi + \phi$$

Thus, $$\hat{P}Y_{lm}(\theta,\phi) = Y_{lm}(\pi-\theta,\pi+\phi).$$

Now, $$e^{im(\pi+\phi)} = e^{im\pi}e^{im\phi} = (-1)^m e^{im\phi},$$

$$P_l^m(\cos(\pi-\theta)) = P_l^m(-\cos\theta)$$

$$= (-1)^{l+m}P_l^m(\cos\theta), \qquad \text{from Eq. (E.38)}$$

Hence, from Eq. (5.55b),

$$Y_{lm}(\pi-\theta,\pi+\phi)=(-1)^l Y_{lm}(\theta,\phi).$$

Addition theorem[7] :

$$\frac{4\pi}{2l+1}\sum_{m=-l}^{+l} Y^*_{lm}(\theta_1,\phi_1)Y_{lm}(\theta_2,\phi_2)=P_l(\cos\theta), \tag{5.62}$$

where, the angles involved are shown in Fig. 5.3.

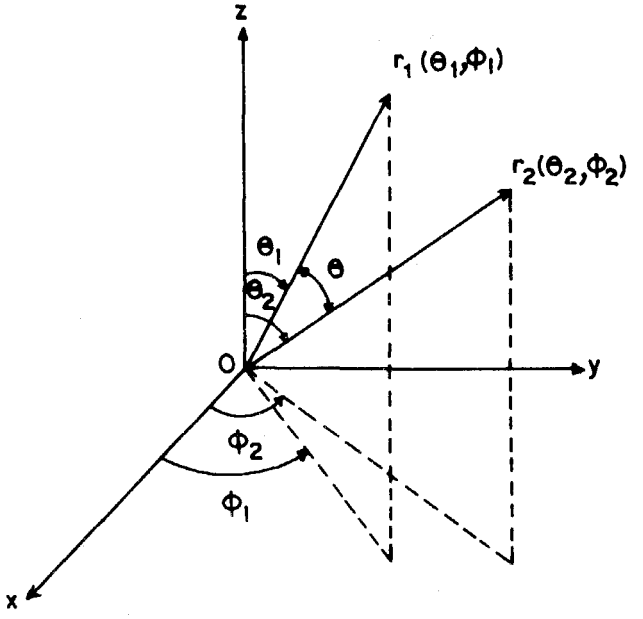


Fig. 5.3. Angles involved in the spherical harmonic addition theorem [Eq. (5.62)].

From the figure, we have,

$$\cos\theta=\frac{\mathbf{r}_1\cdot\mathbf{r}_2}{r_1r_2}=\frac{x_1x_2+y_1y_2+z_1z_2}{r_1r_2}$$

Using, $x_1=r_1\sin\theta_1\cos\phi_1$, etc. [see Eq. (4.107)], we get,

$$\cos\theta=\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\phi_1-\phi_2). \tag{5.62a}$$

The spherical harmonics for the lowest few values of l, are listed below:

$$Y_{00}(\theta,\phi)=\frac{1}{\sqrt{4\pi}},$$

$$Y_{10}(\theta,\phi)=\sqrt{\frac{3}{4\pi}}\cos\theta,$$

7. A derivation of the theorem is given in Section 5.6 [(Eq. (5.152)].

$$Y_{1\pm1}(\theta,\phi) = \mp\sqrt{\frac{3}{8\pi}}\sin\theta e^{\pm i\phi},$$

$$Y_{20}(\theta,\phi) = \sqrt{\frac{5}{16\pi}}(3\cos^2\theta - 1),$$

$$Y_{2\pm1}(\theta,\phi) = \mp\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{\pm i\phi},$$

$$Y_{2\pm2}(\theta,\phi) = \sqrt{\frac{15}{32\pi}}\sin^2\theta e^{\pm 2i\phi},$$

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

$$Y_{l\pm l}(\theta,\phi) = c_{\pm l}(\sin\theta)^l e^{\pm il\phi}, \tag{5.63}$$

where $c_{\pm l}$ is independent of θ and ϕ.

Since the dependence of $Y_{lm}(\theta,\phi)$ on ϕ is contained in the factor $e^{\pm im\phi}$, the absolute value $|Y_{lm}(\theta,\phi)|$ is independent of ϕ. In Fig. 5.4, we have plotted $|Y_{lm}(\theta,\phi)|$ as a function of θ, for the lowest three values of l, where $|Y_{lm}(\theta,\phi)|$ is proportional to the radial distance from the centre. Note that there is only one value of $|Y_{lm}(\theta,\phi)|$ for a given value of θ.

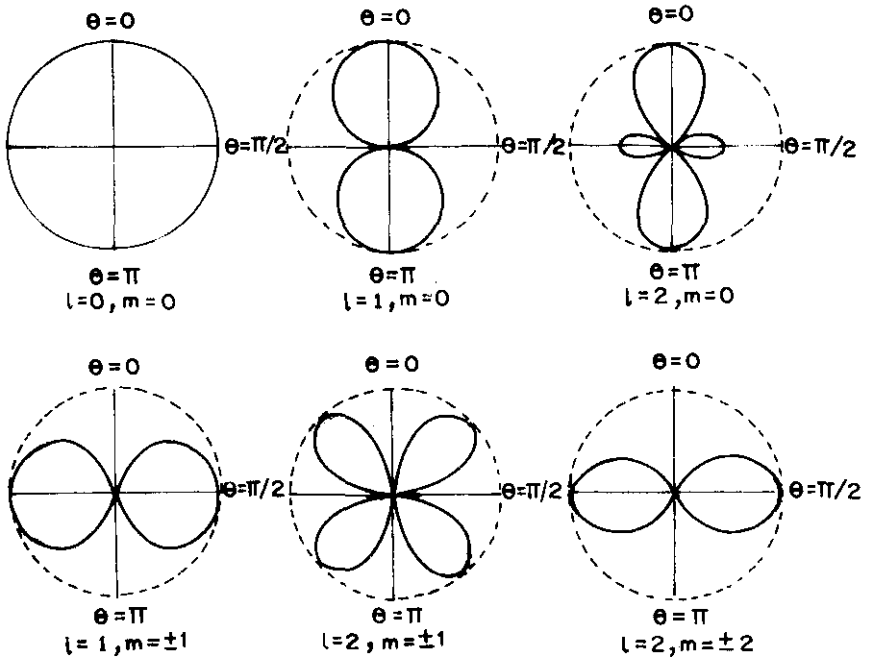


Fig. 5.4. Polar diagram showing $|Y_{lm}(\theta,\phi)|$ as a function of θ.

Now, the probability density for the particle to be found at θ is given by

$$\mathcal{P}_{lm}(\theta) = \int_0^{2\pi} | Y_{lm}(\theta, \phi) |^2 \, d\phi = 2\pi | Y_{lm}(\theta, \phi) |^2 . \tag{5.64}$$

Let us consider the two extreme cases, $m = 0$ and $m = \pm l$.

m = ±l

In this case, the angular momentum vector is 'parallel' to the z-axis so that the particle-orbit should be in the xy-plane ($\theta = \pi/2$). We see, from Fig. 5.5, that this classical expectation is not fulfilled except for very large values of l. For low values of l, there is appreciable probability for the particle-orbit to be at an angle to the xy-plane.[8]

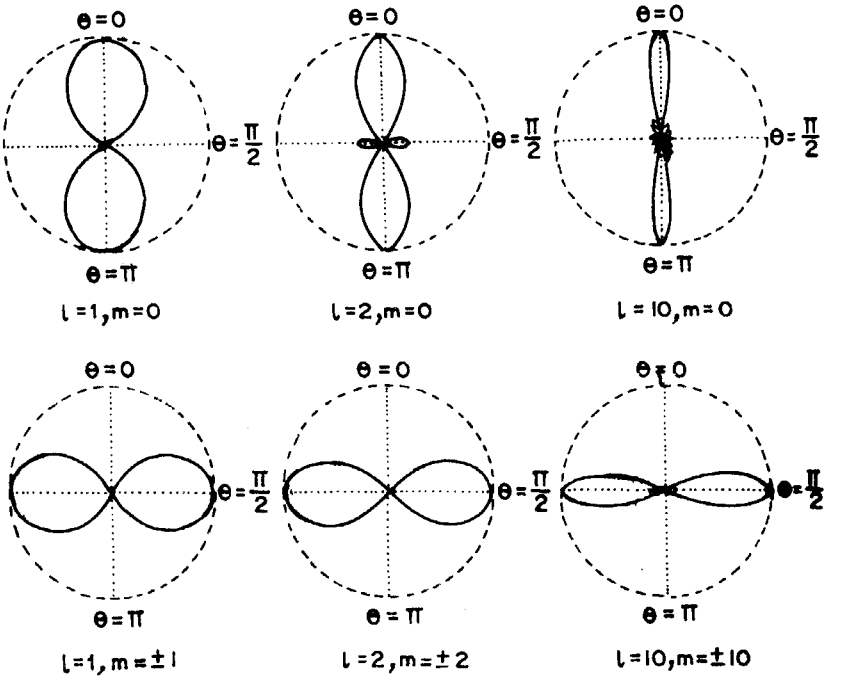


Fig. 5.5. Polar diagram of $\mathcal{P}_{lm}(\theta)$ given by Eq. (5.64).

m = 0

The angular momentum vector should be in the xy-plane, so that the plane of the particle orbit should make an angle zero with the z-axis. Again, this expectation is realized only in the limit of very large angular momenta.[9]

---

8. Note that the plane of the orbit undergoes a continuous change corresponding to the precessional motion of the angular momentum vector about the z-axis (see Section 5.8).
9. For further discussion and diagrams, see, Pauling, L. and Wilson, E. B. *Introduction to Quantum Mechanics* (McGraw-Hill). International Student Edition (Kogakusha Co., Tokyo), Section 21d.

5.5 ADDITION OF ANGULAR MOMENTA

5.5A Clebsch-Gordon Coefficients

Angular momenta, being vectors, could be added vectorially. Thus, for a system consisting of two subsystems with angular momenta $\mathbf{J}_1$ and $\mathbf{J}_2$, we can define a resultant (or total) angular momentum $\mathbf{J}$ by (see Fig. 5.6),

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 = \mathbf{J}_2 + \mathbf{J}_1 \tag{5.65}$$

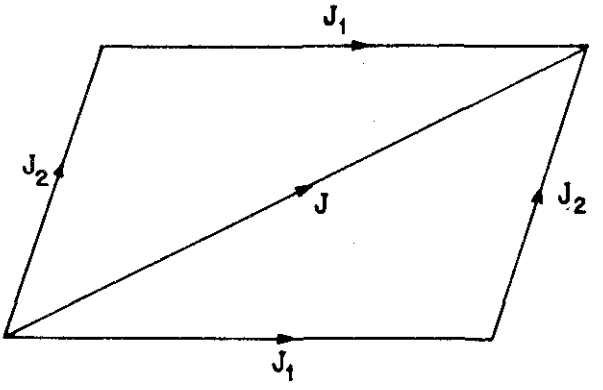


Fig. 5.6 Vectorial addition of two angular momenta.

Corresponding to the three angular momentum vectors $\mathbf{J}_1, \mathbf{J}_2$ and $\mathbf{J}$, we have the six Hermitian operators $\hat{\mathbf{J}}_1^2, \hat{J}_{1z}, \hat{\mathbf{J}}_2^2, \hat{J}_{2z}, \hat{\mathbf{J}}^2$ and $\hat{J}_z$. All these six operators, however, do not commute among themselves ($\hat{J}_{1z}$ and $\hat{J}_{2z}$ do not commute with $\hat{\mathbf{J}}^2$). But we can form two sets consisting of four operators each, which, together with the Hamiltonian, form complete sets of commuting observables for the system.

These are:

(i) $\hat{\mathbf{J}}_1^2, \hat{\mathbf{J}}_2^2, \hat{J}_{1z}, \hat{J}_{2z}$

and (ii) $\hat{\mathbf{J}}_1^2, \hat{\mathbf{J}}_2^2, \hat{\mathbf{J}}^2, \hat{J}_z$.

The basis vectors defined by set (i) will be denoted[10] by

$$|j_1 j_2 m_1 m_2\rangle \equiv |j_1 m_1\rangle\, |j_2 m_2\rangle,$$

or, briefly, by $|m_1 m_2\rangle$, while those defined by set (ii) will be denoted by $|j_1 j_2 jm\rangle$, or $|jm\rangle$. Thus,

---

10. The quantum number α, specifying the eigenvalues of the Hamiltonian, is suppressed. Thus, $|j_1 j_2 m_1 m_2 \equiv |\alpha j_1 j_2 m_1 m_2>$.

$$\hat{\mathbf{J}}_i^2 \mid j_1 j_2 m_1 m_2\rangle = j_i(j_i+1)\hbar^2 \mid j_1 j_2 m_1 m_2\rangle. \tag{5.66a}$$

$$\hat{J}_{iz} \mid j_1 j_2 m_1 m_2\rangle = m_i\hbar \mid j_1 j_2 m_1 m_2\rangle, \tag{5.66b}$$

$$\hat{\mathbf{J}}_i^2 \mid j_1 j_2 jm\rangle = j_i(j_i+1)\hbar^2 \mid j_1 j_2 jm\rangle, \tag{5.67a}$$

$$\hat{\mathbf{J}}^2 \mid j_1 j_2 jm\rangle = j(j+1)\hbar^2 \mid j_1 j_2 jm\rangle, \tag{5.67b}$$

$$\hat{\mathbf{J}}_z \mid j_1 j_2 jm\rangle = m\hbar \mid j_1 j_2 jm\rangle. \tag{5.67c}$$

The representation defined by the vectors $\mid m_1 m_2\rangle$ is called the *uncoupled representation* whereas the vectors $\mid jm\rangle$ define the *coupled representation*. Since there are $(2j_1+1)$ different values of m_1, for a given value of m_2 and $(2j_2+1)$ values of m_2 for each value of m_1, the dimensionality of the representation is $(2j_1+1)(2j_2+1)$.

Now, since $\{\mid jm\rangle\}$ and $\{\mid m_1 m_2\rangle\}$ are merely different bases in the same Hilbert space, they should be related by a Unitary transformation. Thus (see Eq. (2.124b)),

$$\mid jm\rangle = \sum_{m_1,m_2} U^{j_1 j_2}_{m_1 m_2; jm} \mid m_1 m_2\rangle, \tag{5.68}$$

where $U^{j_1 j_2}_{ij}$ is the ijth element of the unitary matrix $U^{j_1 j_2}$ that transforms the basis $\{\mid m_1 m_2 >\}$ to the basis $\{\mid jm\rangle\}$.

Using the closure property of the basis $\{\mid m_1 m_2\rangle\}$, we can write [see Eq. (2.27[1])]

$$\mid jm\rangle = \sum_{m_1 m_2} \mid j_1 j_2 m_1 m_2\rangle \langle j_1 j_2 m_1 m_2 \mid jm\rangle. \tag{5.69}$$

Comparing Eqs. (5.68) and (5.69), we see that,

$$U^{j_1 j_2}_{m_1 m_2; jm} \equiv \langle j_1 j_2 m_1 m_2 \mid jm\rangle. \tag{5.70}$$

The R.H.S. of Eq. (5.70) should actually read: $\langle \alpha j_1 j_2 m_1 m_2 \mid \alpha jm\rangle$ (see footnote 10). However, since $\{\mid \alpha jm\rangle\}$ and $\{\mid \alpha j_1 j_2 m_1 m_2 >\}$ differ only with regard to their 'orientations' relative to the angular momentum vectors, the transforming matrix $U^{j_1 j_2}$ should be independent of α. Thus, the matrix element (5.70) depends only on the six angular momentum quantum numbers. Various symbols and names are used in the literature for this matrix element. We will adopt the symbol.

$C^{j_1 j_2 j}_{m_1 m_2 m}$ and the names, *Clebsch-Gordon*, or *C-coefficient*.

Thus,

$$C^{j_1 j_2 j}_{m_1 m_2 m} \equiv U^{j_1 j_2}_{m_1 m_2; jm} \tag{5.71}$$

Eq. (5.68) becomes:

$$\mid jm\rangle = \sum_{m_1 m_2} C^{j_1 j_2 j}_{m_1 m_2 m} \mid j_1 m_1\rangle \mid j_2 m_2\rangle. \tag{5.68a}$$

This equation could be regarded as the defining equation for the C-coefficients.

The Selection Rules

For given values of j_1, j_2, m_1 and m_2, $C^{j_1 j_2 j}_{m_1 m_2 m}$ is non-zero only for certain allowed values of j and m. The rules specifying these allowed values are referred to as *selection rules*. These are:

(SR 1). $m = m_1 + m_2$; (5.72)

(SR 2). $|j_1 - j_2| \leq j \leq (j_1 + j_2)$, where, j varies by integer steps.

(SR 1) follows from the relation, $\hat{J}_z = \hat{J}_{1z} + \hat{J}_{2z}$.

For,

$$\hat{J}_z \mid jm > = (\hat{J}_{1z} + \hat{J}_{2z}) \mid jm >$$

$$= \sum_{m_1 m_2} C^{j_1 j_2 j}_{m_1 m_2 m} (\hat{J}_{1z} + \hat{J}_{2z}) \mid j_1 j_2 m_1 m_2 >$$

$$= \sum_{m_1 m_2} (m_1 + m_2)\hbar\, C^{j_1 j_2 j}_{m_1 m_2 m} \mid j_1 j_2 m_1 m_2 > .$$

i.e., $$\sum_{m_1 m_2} (m - m_1 - m_2) C^{j_1 j_2 j}_{m_1 m_2 m} \mid j_1 j_2 m_1 m_2 > = 0.$$

Since the vectors $\mid j_1 j_2 m_1 m_2 >$ are linearly-independent, this implies that,

either, $$C^{j_1 j_2 j}_{m_1 m_2 m} = 0, \; m \neq (m_1 + m_2),$$

or $$C^{j_1 j_2 j}_{m_1 m_2 m} \neq 0, \; m = (m_1 + m_2).$$

(SR 2) may be derived as follows:

Maximum value of $m = m_{max} \equiv (m_1 + m_2)_{max} = (j_1 + j_2)$. Thus, maximum value of $j \equiv j_{max} = m_{max} = (j_1 + j_2)$. The next lower value of $m = j_1 + j_2 - 1$. There are two states with this value of m, namely, $\mid m_1 m_2) \equiv \mid j_1 j_2 - 1)$ and $\mid j_1 - 1 j_2)$. One of these belongs to $j = (j_1 + j_2)$ and the other to $j = j_1 + j_2 - 1$. Similarly, there are three states with $m = j_1 + j_2 - 2$, the corresponding j-values being $(j_1 + j_2), (j_1 + j_2 - 1)$ and $(j_1 + j_2 - 2)$. In general, there are $(p + 1)$ states with $m = j_1 + j_2 - p$, the j-values being $(j_1 + j_2), (j_1 + j_2 - 1), \ldots, (j_1 + j_2 - p)$. The maximum value of p is given by

$$(j_2 - p_{max}) = (m_2)_{min} = -j_2, \text{ if } j_2 < j_1,$$

and by $$(j_1 - p_{max}) = (m_1)_{min} = -j_1, \text{ if } j_1 < j_2$$

That is,

$$p_{max} = 2_{j<}, \text{ where } j_< \text{ is the lesser } j_i.$$

Thus,
$$j_{\min} = j_1 + j_2 - p_{\max}$$
$$= (j_1 - j_2), \text{ if } j_1 > j_2$$
$$= (j_2 - j_1), \text{ if } j_2 > j_1$$

or,
$$j_{\min} = |j_1 - j_2|, \tag{5.73a}$$

and
$$j = |j_1 - j_2|, |j_1 - j_2| + 1, \ldots, (j_1 + j_2). \tag{5.73b}$$

This method of obtaining the allowed values of j is illustrated in Table 5.1 for the case $j_1 = 3/2, j_2 = 5/2$.

Table 5.1

j-values corresponding to $j_1 = 3/2$, j_2, 5/2. ($p_{\max} = 2j_1 = 3$)

m	m_1	m_2	j
$j_1 + j_2 = 4$	$j_1 = 3/2$	$j_2 = 5/2$	$m = 4$
$j_1 + j_2 - 1 = 3$	$j_1 = 3/2$	$j_2 - 1 = 3/2$	$j_1 + j_2 = 4;$
	$j_1 - 1 = 1/2$	$j_2 = 5/2$	$j_1 + j_2 - 1 = 3.$
$j_1 + j_2 - 2 = 2$	$j_1 = 3/2$	$j_2 - 2 = 1/2$	$j_1 + j_2 = 4;$
	$j_1 - 1 = 1/2$	$j_2 - 1 = 3/2$	$j_1 + j_2 - 1 = 3;$
	$j_1 - 2 = -1/2$	$j_2 = 5/2$	$j_1 + j_2 - 2 = 2.$
$j_1 + j_2 - 3 = 1$	$j_1 = 3/2$	$j_2 - 3 = -1/2$	$j_1 + j_2 = 4;$
	$j_1 - 1 = 1/2$	$j_2 - 2 = 1/2$	$j_1 + j_2 - 1 = 3;$
	$j_1 - 2 = -1/2$	$j_2 - 1 = 3/2$	$j_1 + j_2 - 2 = 2;$
	$j_1 - 3 = -3/2$	$j_2 = 5/2$	$j_1 + j_2 - 3 = 1.$

The matrix $U^{j_1 j_2}$ is shown below:

		$i \to j_1+j_2$	j_1+j_2	j_1+j_2-1	.	.	.	. .	j_1+j_2
		$m \to j_1+j_2$	j_1+j_2-1	j_1+j_2-1	.	.	.	. .	$-j_1-j_2$
m_1 ↓	m_2 ↓								
j_1	j_2	C_{11}	0	0	.	.	.	. . .	0
j_1-1	j_2	0	C_{22}	C_{23}	0	.	.	. . .	0
j_1	j_2-1	0	C_{32}	C_{33}	0	.	.	. . .	0
j_1-2	j_2	0	0	0	C_{44}	C_{45}	C_{46}	. . .	0
j_1-1	j_2-1	0	0	0	C_{54}	C_{55}	C_{56}	. . .	0
j_1	j_2-2	0	0	0	C_{64}	C_{65}	C_{66}	. . .	0
•		.	.	.	.	.	.	. . .	0
•		.	.	.	.	.	.	. .	0
•		.	.	.	.	.	.	. .	0
$-j_1$	$-j_2$	0	0	0	0	0	0	. . .	C_{ii}

Example: $j_1 = 3/2, j_2 = 1/2$.

		$j \to 2$	2	1	2	1	2	1	2
		$m \to 2$	1	1	0	0	−1	−1	−2
m_1 ↓	m_2 ↓								
3/2	1/2	(C_{11})	0	0	0	0	0	0	0
1/2	1/2	0	C_{22}	C_{23}	0	0	0	0	0
3/2	−1/2	0	C_{32}	C_{33}	0	0	0	0	0
−1/2	1/2	0	0	0	C_{44}	C_{45}	0	0	0
1/2	−1/2	0	0	0	C_{54}	C_{55}	0	0	0
−3/2	1/2	0	0	0	0	0	C_{66}	C_{67}	0
−1/2	−1/2	0	0	0	0	0	C_{76}	C_{77}	0
−3/2	−1/2	0	0	0	0	0	0	0	(C_{88})

Problem 5.10: Verify that the number of independent vectors $|jm\rangle$ for given j_1 and j_2 is $(2j_1+1)(2j_2+1)$.

Choice of Phase

In Eq. (5.68a) for a given $|j_1m_1\rangle$ and $|j_2m_2\rangle$, the phase of $|jm\rangle$ could be chosen arbitrarily. Correspondingly, there is a certain amount of arbitrariness in the phase of the C-coefficient. The adoption of some convention in fixing the phase is, therefore, necessary to avoid confusion. The following is the procedure conventionally adopted :

For $j = j_1 + j_2$ and $m = j_1 + j_2$, Eq. (5.68a) reduces to

$$|j_1+j_2 j_1+j_2\rangle = C^{j_1 j_2 j_1+j_2}_{j_1 j_2 j_1+j_2} |j_1 j_1\rangle |j_2 j_2\rangle \tag{5.74}$$

From the orthonormality of the wavefunctions involved here, we have,

$$\langle j_1+j_2 j_1+j_2 | j_1+j_2 j_1+j_2\rangle = \left| C^{j_1 j_2 j_1+j_2}_{j_1 j_2 j_1+j_2} \right|^2 = 1.$$

Therefore, $C^{j_1 j_2 j_1+j_2}_{j_1 j_2 j_1+j_2} = e^{i\delta}$, where δ is real.

Now, we choose the phase of $|j_1+j_2 j_1+j_2\rangle$ in Eq. (5.74) such that $\delta = 0$. That is,

$$C^{j_1 j_2 j_1+j_2}_{j_1 j_2 j_1+j_2} = 1. \tag{5.75}$$

This choice fixes the phases of all the eigenvectors with $j = j_1 + j_2$, according to Eq. (5.31b), and therefore it fixes also the phases of all the C-coefficients belonging to $j = j_1 + j_2$.

Problem 5.11: Verify the above statement for the case $j = j_1 + j_2$ and $m = j_1 + j_2 - 1$.

Now, Eqs. (5.31a, b) do not enable us to fix the relative phases of eigenvectors belonging to different j-values. This is because the operators $\hat{J}_+$ and $\hat{J}_-$ do not have non-zero matrix elements between such eigenvectors. The remedy, therefore, lies in finding an operator which does have non-vanishing matrix elements between vectors belonging to different j-values. One such operator is $\hat{\mathbf{J}}_1$ (or $\hat{\mathbf{J}}_2$). The following commutation relations could be deduced from the commutation relations (5.7a) and (5.8):

$$[\hat{J}_z, \hat{J}_{1z}] = \hat{0} \tag{5.76a}$$

$$[\hat{J}_z, \hat{J}_{1\pm}] = \pm\hbar \hat{J}_{1\pm} \tag{5.76b}$$

$$[\hat{J}_\pm, \hat{J}_{1z}] = \mp\hbar \hat{J}_{1\pm} \tag{5.76c}$$

$$[\hat{J}_\pm, \hat{J}_{1\pm}] = \hat{0} \tag{5.76d}$$

$$[\hat{J}_\pm, \hat{J}_{1\mp}] = \pm 2\hbar \hat{J}_{1z}. \tag{5.76e}$$

Comparing these with the commutation relations (5.16 a-c), we see that these are equivalent to the equation [see Eq. (5.6b)],

$$\hat{\mathbf{J}} \times \hat{\mathbf{J}}_1 = i\hbar \hat{\mathbf{J}}_1. \tag{5.76}$$

The following relationships follow from Eqs. (5.76) and (5.65):

$$\hat{\mathbf{J}} \times \hat{\mathbf{J}}_1 = -\hat{\mathbf{J}}_1 \times \hat{\mathbf{J}} + 2i\hbar \hat{\mathbf{J}}_1. \tag{5.77a}$$

$$\hat{\mathbf{J}} \cdot \hat{\mathbf{J}}_1 = \hat{\mathbf{J}}_1 \cdot \hat{\mathbf{J}} = \frac{1}{2}(\hat{\mathbf{J}}^2 + \hat{\mathbf{J}}_1^2 - \hat{\mathbf{J}}_2^2). \tag{5.77b}$$

$$[\hat{\mathbf{J}}, (\hat{\mathbf{J}} \cdot \hat{\mathbf{J}}_1)] = \hat{0}. \tag{5.77c}$$

$$[\hat{\mathbf{J}}^2, \hat{\mathbf{J}}_1] = -i\hbar(\hat{\mathbf{J}} \times \hat{\mathbf{J}}_1 - \hat{\mathbf{J}}_1 \times \hat{\mathbf{J}}) \tag{5.77d}$$

$$= -2i\hbar(\hat{\mathbf{J}} \times \hat{\mathbf{J}}_1 - i\hbar \hat{\mathbf{J}}_1) \tag{5.77e}$$

$$[\hat{\mathbf{J}}^2, [\hat{\mathbf{J}}^2, \hat{\mathbf{J}}_1]] \equiv \hat{\mathbf{J}}^4 \hat{\mathbf{J}}_1 - 2\hat{\mathbf{J}}^2 \hat{\mathbf{J}}_1 \hat{\mathbf{J}}^2 + \hat{\mathbf{J}}_1 \hat{\mathbf{J}}^4$$

$$= 2\hbar^2(\hat{\mathbf{J}}^2 \hat{\mathbf{J}}_1 + \hat{\mathbf{J}}_1 \hat{\mathbf{J}}^2) - 4\hbar^2 \hat{\mathbf{J}}(\hat{\mathbf{J}} \cdot \hat{\mathbf{J}}_1). \tag{5.77f}$$

From Eqs. (5.76a, b) and (5.77f), the following selection rules for the matrix elements of $\hat{\mathbf{J}}_1$ follow:

$$\langle j'm' \mid \hat{J}_{1z} \mid jm \rangle = 0, \text{ unless } m' = m; (j' - j) = 0, \pm 1. \tag{5.78a}$$

$$\langle j'm' \mid \hat{J}_{1\pm} \mid jm \rangle = 0, \text{ unless } m' = m \pm 1; (j' - j) = 0, \pm 1. \tag{5.78b}$$

Problem 5.12: Deduce the selection rules (5.78a, b).

From Eqs. (5.66b) and (5.68a), we have,

$$\langle j'm \mid \hat{J}_{1z} \mid jm \rangle = \hbar \sum_{m_1, m_2} m_1 C^{j_1 j_2 j}_{m_1 m_2 m} C^{j_1 j_2 j'}_{m_1 m_2 m}. \tag{5.79}$$

Also,

$$\langle j'm \pm 1 \mid \hat{J}_{1\pm} \mid jm \rangle = \hbar \sum_{m_1, m_2} \sqrt{(j_1 \mp m_1)(j_1 \pm m_1 + 1)} \times C^{j_1 j_2 j}_{m_1 m_2 m} C^{j_1 j_2 j'}_{m_1 \pm 1 m_2 m \pm 1} \tag{5.80}$$

In Eqs. (5.79) and (5.80), $j' = j$ or $j' = j \pm 1$. Thus, knowledge of the matrix elements on the L.H.S. together with a knowledge of the phases of $C^{j_1 j_2 j}_{m_1 m_2 m}$ would enable us to fix the phases of $C^{j_1 j_2 j \pm 1}_{m'_1 m'_2 m'}$. The following convention is adopted for the phases of the matrix elements:

(i) $$\langle j \pm 1, m \mid \hat{J}_{1z} \mid jm \rangle = \text{ real and positive.} \tag{5.81a}$$

This implies, since $\langle j \pm 1, m \mid (\hat{J}_{1z} + \hat{J}_{2z}) \mid jm \rangle = 0$, that

$$\langle j \pm 1, m \mid \hat{J}_{2z} \mid jm \rangle = \text{ real and negative.}$$

(ii) $$\langle J \pm 1, m + 1 \mid \hat{J}_{1+} \mid jm \rangle = \text{ real and positive.} \tag{5.81b}$$

These conventions result in the reality of all Clebsch-Gordon coefficients.

Racah[11] gives the following explicit expression for the C-coefficient:

$$C^{j_1 j_2 j}_{m_1 m_2 m} = \delta_{m_1+m_2,m}\left[\frac{(2j+1)(j_1+j_2-j)!(j+j_1-j_2)!(j+j_2-j_1)!}{(j_1+j_2+j+1)!}\right]^{1/2}$$

$$\times\sum_{\kappa}(-1)^{\kappa}\frac{\{(j+m)!(j-m)!(j_1+m_1)!(j_1-m_1)!(j_2+m_2)!(j_2-m_2)!\}^{1/2}}{\kappa!(j_1+j_2-j-\kappa)!(j_1-m_1-\kappa)!(j_2+m_2-\kappa)!(j-j_2+m_1+\kappa)!}$$

$$\times\frac{1}{(j-j_1-m_2+\kappa)!} \tag{5.82}$$

Here κ takes all integer values consistent with the factorial notation (Factorial of a negative number is infinite).

Problem 5.13: Denote the coupled state of two spin $\frac{1}{2}$ particles by $|SM\rangle$. Show that the eigenvalues of the operator $(\hat{\sigma}_1 \cdot \hat{\sigma}_2)$ are -3 and $+1$ in the single $t(S=0)$ and the triplet $(S=1)$ states, respectively. Hence, construct the projection operators $\hat{\pi}_s$ and $\hat{\pi}_t$ for the single t and the triplet states.

Properties of the C-Coefficients:

Symmetry:

$$C^{j_1 j_2 j}_{m_1 m_2 m} = (-1)^{j_1+j_2-j}C^{j_1 j_2 j}_{-m_1,-m_2,-m} \tag{5.83a}$$

$$= (-1)^{j_1+j_2-j}C^{j_2 j_1 j}_{m_2 m_1 m} \tag{5.83b}$$

$$= (-1)^{j_1-m_1}\left[\frac{2j+1}{2j_2+1}\right]^{1/2}C^{j_1 j j_2}_{m_1,-m,-m_2} \tag{5.83c}$$

Other symmetry relations could be obtained by the application of one or more of the above relations.[12] These symmetry relations follow from the expression (5.82) for the C-coefficient.

Orthogonality:

$$\sum_{m_1 m_2} C^{j_1 j_2 j}_{m_1 m_2 m}C^{j_1 j_2 j'}_{m_1 m_2 m'} = \delta_{jj'}\delta_{mm'} \tag{5.84a}$$

$$\sum_{j} C^{j_1 j_2 j}_{m_1 m_2 m}C^{j_1 j_2 j}_{m'_1 m'_2 m} = \delta_{m_1 m'_1}\delta_{m_2 m'_2} \tag{5.84b}$$

---

11. Racah, G. *Physical Review*, **62**, 438 (1942).
12. An important one is:

$$C^{j_1 j_2 j}_{m_1 m_2 m} = (-1)^{j_2+m_2}\left(\frac{2j+1}{2j_1+1}\right)^{1/2}C^{j j_2 j_1}_{-m m_2 -m_1}$$

$$\sum_{m_2} C^{j_1 j_2 j}_{m_1 m_2 m} C^{j'_1 j_2 j}_{m_1 m_2 m} = \left(\frac{2j+1}{2j_1+1}\right)\delta_{j_1 j'_1} \tag{5.84c}$$

$$\sum_{m_1} C^{j_1 j_2 j}_{m_1 m_2 m} C^{j_1 j'_2 j}_{m_1 m_2 m} = \left(\frac{2j+1}{2j_2+1}\right)\delta_{j'_2 j_2} \tag{5.84d}$$

In Eqs. (5.84a, b), $m = m_1 + m_2$ is fixed whereas in (5.84c) $m_1 = m - m_2$ and in (5.84d), $m_2 = m - m_1$ are fixed. Relations (5.84a, b) follow from the unitarity of the matrix $U^{j_1 j_2}$ [see Eqs. (5.71) and (2.120a, b)]. (5.84c) and (5.84d) could be obtained by writing,

$$|j_1 m_1\rangle = \sum_{m_2 m} C^{j_2 j j_1}_{-m_2 m m_1} |j_2 - m_2\rangle |jm\rangle,$$

$$|j_2 m_2\rangle = \sum_{m_1 m} C^{j_1 j j_2}_{-m_1 m m_2} |j_1 - m_1\rangle |jm\rangle,$$

and then, using the orthogonality relationships

$$\langle j'_1 m_1 | j_1 m_1\rangle = \delta_{j_1 j'_1},$$

$$\langle j'_2 m_2 | j_2 m_2\rangle = \delta_{j_2 j'_2},$$

and the symmetry relationships (5.83c, d).

Recursion Relations:

$$[(j-m)(j+m+1)]^{1/2} C^{j_1 j_2 j}_{m_1 m_2 m+1}$$

$$= [(j_1+m_1)(j_1-m_1+1)]^{1/2} C^{j_1 j_2 j}_{m_1-1 m_2 m}$$

$$+[(j_2+m_2)(j_2-m_2+1)]^{1/2} C^{j_1 j_2 j}_{m_1 m_2-1 m} \tag{5.85a}$$

$$[(j+m)(j-m+1)]^{1/2} C^{j_1 j_2 j}_{m_1 m_2 m-1}$$

$$= [(j_1-m_1)(j_1+m_1+1)]^{1/2} C^{j_1 j_2 j}_{m_1+1 m_2 m}$$

$$\times[(j_2-m_2)(j_2+m_2+1)]^{1/2} C^{j_1 j_2 j}_{m_1 m_2+1 m} \tag{5.85b}$$

$$\{(j-m)(j+m)\}^{1/2} A_{j-1,j}(j_1 j_2) C^{j_1 j_2 j-1}_{m_1 m_2 m}$$

$$= \{m_1 - m A_{jj}(j_1 j_2)\} C^{j_1 j_2 j}_{m_1 m_2 m} - \{(j-m+1)(j+m+1)\}^{1/2}$$

$$\times A_{j+1,j}(j_1 j_2) C^{j_1 j_2 j+1}_{m_1 m_2 m} \tag{5.85c}$$

where

$$A_{j'j}(j_1 j_2) = \left[\frac{(j'+j_2-j_1+1)(j'+j_1-j_2+1)(j_1+j_2+j+1)(j_1+j_2-j+1)}{4j^2(2j+1)(2j'+1)} \right]^{1/2} \tag{5.86}$$

Relation (5.85a) is obtained by operating on $|jm>$ with $\hat{J}_+$ and equating the result to $(\hat{J}_{1+}+\hat{J}_{2+})|jm>$, while (5.85b) is obtained by an identical procedure, employing the operator $\hat{J}_- = \hat{J}_{1-}+\hat{J}_{2-}$. (5.85c) is based on the matrix elements of the operator $\hat{\mathbf{J}}_1$. The explicit formula (5.82) for C-coefficient is, in fact, derived from these recursion relations (see, reference quoted in footnote 11).

Problem 5.14: Using the symmetry relations, show that:

(a) $C_{000}^{l_1 l_1 l} = 0$, unless $l_1 + l_2 + l$ is even.

(b) $C_{11\,\frac{1}{2}\frac{1}{2}1}^{j_1 j_2 j} = 0$, unless j is odd.

Problem 5.15: Evaluate the Clebsch-Gordon coefficients involved in the angular momentum coupling of two spin-half particles.

Other Related Coefficients :

$$D_{l_1 l_2 l} = \left\{ \frac{(2l_1+1)(2l_2+1)}{(2l+l)} \right\}^{1/2} C_{000}^{l_1 l_2 l}. \tag{5.87}$$

$$V(j_1 j_2 j_3; m_1 m_2 m_3) = \frac{1}{\sqrt{2j_3+1}} (-1)^{j_3 - m_3} C_{m_1 m_2 - m_3}^{j_1 j_2 j_3} \tag{5.88}$$

$$= S_{j_1 m_1,\, j_2 m_2,\, j_3 m_3} \tag{5.89}$$

3j-symbol:

$$\begin{pmatrix} j_1 j_2 j_3 \\ m_1 m_2 m_3 \end{pmatrix} = \frac{1}{\sqrt{2j_3+1}} (-1)^{j_1 - j_2 - m_3} C_{m_1 m_2 - m_3}^{j_1 j_2 j_3} \tag{5.90}$$

$$= X(j_1 j_2 j_3; m_1 m_2 m_3). \tag{5.91}$$

The V-coefficient is due to Racah, the S-coefficient due to Fano, the X-coefficient due to Schwinger and the $3j$-symbol due to Wigner. The $3j$-symbol and the C-coefficients are the ones most widely used.

5.5B Racah Coefficients

When the system is composed of three subsystems with angular momenta $\mathbf{J}_1, \mathbf{J}_2$ and $\mathbf{J}_3$, the total angular momentum $\mathbf{J}$ is given by

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 + \mathbf{J}_3. \tag{5.92}$$

The quantum mechanical problem of angular momentum addition now consists in obtaining the wavefunction $|JM\rangle$ of the system in the coupled representation, in terms of the $(2j_1+1)(2j_2+1)(2j_3+1)$ basis vectors, $|j_1j_2j_3m_1m_2m_3\rangle \equiv |j_1m_1\rangle\,|j_2m_2\rangle\,|j_2m_3\rangle$, of the uncoupled representation, where,

$$\left.\begin{aligned}\hat{\mathbf{J}}_i^2\,|j_im_i\rangle &= j_i(j_i+1)\hbar^2\,|j_im_i\rangle\\ \hat{\mathbf{J}}_{iz}\,|j_im_i\rangle &= m_i\hbar\,|j_im_i\rangle\end{aligned}\right\}\ i=1,2,3. \tag{5.93}$$

and

$$\hat{\mathbf{J}}^2\,|JM\rangle = J(J+1)\hbar^2\,|JM\rangle, \tag{5.94a}$$

$$\hat{J}_z\,|JM\rangle = M\hbar\,|JM\rangle. \tag{5.94b}$$

However, in this case, the expression of $|JM\rangle$ in terms of the $|j_1j_2j_3m_1m_2m_3\rangle$, is not unique; for, there are three different ways in which we can couple the individual angular momenta to obtain the resultant:

(a) Add $\mathbf{J}_1$ and $\mathbf{J}_2$ to obtain $\mathbf{J}_{12}$ and then, add $\mathbf{J}_{12}$ and $\mathbf{J}_3$ to obtain $\mathbf{J}$. In this case[13],

$$|JM\rangle \equiv |j_1j_2(J_{12}),j_3;JM\rangle$$

$$= \sum_{m_1m_2m_3} C^{j_1j_2J_{12}}_{m_1m_2M_{12}} C^{J_{12}j_3J}_{M_{12}m_3M}\,|j_1j_2j_3m_1m_2m_3\rangle, \tag{5.95^1}$$

where [see Eq. (5.72)],

$$M_{12} = m_1 + m_2;$$

$$M = M_{12} + m_3 = m_1 + m_2 + m_3 \text{ (fixed)} \tag{5.96}$$

(b) Add $\mathbf{J}_2$ and $\mathbf{J}_3$ to obtain $\mathbf{J}_{23}$ and, then, add $\mathbf{J}_1$ and $\mathbf{J}_{23}$. Correspondingly, we have,

$$|JM\rangle \equiv |j_1,j_2j_3(J_{23});JM\rangle$$

$$= \sum_{m_1,m_2,m_3} C^{j_2j_3J_{23}}_{m_2m_3M_{23}} C^{j_1J_{23}J}_{m_1M_{23}M}\,|j_1j_2j_3m_1m_2m_3\rangle \tag{5.95^2}$$

with $M_{23} = m_2 + m_3$ and $M = m_1 + m_2 + m_3$ (fixed).

(c) Add $\mathbf{J}_1$ and $\mathbf{J}_3$ to obtain $\mathbf{J}_{13}$, then, add $\mathbf{J}_2$ to $\mathbf{J}_{13}$. We have, then,

$$|JM\rangle = |j_1j_3(J_{13}),j_2;JM\rangle$$

$$= \sum_{m_1,m_2,m_3} C^{j_1j_3J_{13}}_{m_1m_3M_{13}} C^{J_{13}j_2J}_{M_{13}m_2M}\,|j_1j_2j_3m_1m_2m_3\rangle \tag{5.95^3}$$

These different ways of coupling the angular momenta are shown schematically in Fig. 5.7.

---

13. As usual, we omit the quantum numbers other than those related to the angular momentum, from the specification of the state.

Now, the right hand sides of Eq. (5.95[1-3]) represent different sets of basis vectors in terms of which the eigenfunctions of the total angular momentum can be expressed. They should, therefore, be related by unitary transformations. Thus [see Eq. (2.124b)],

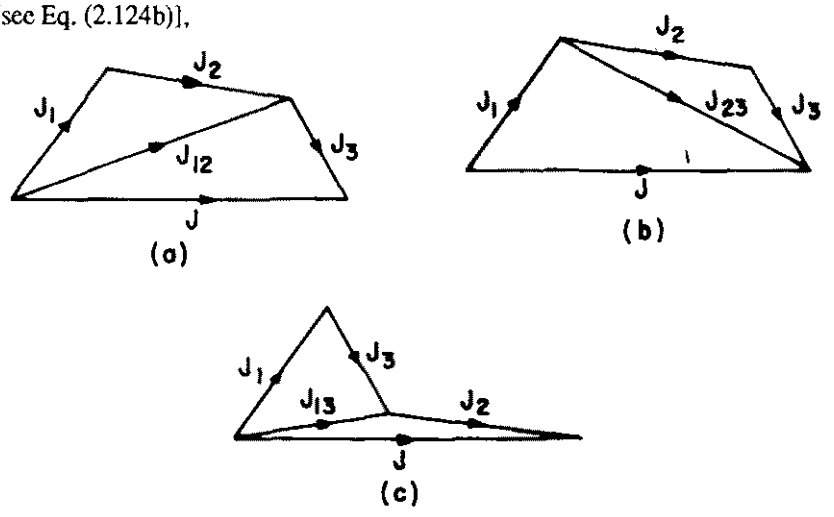


Fig. 5.7 Different ways of coupling three angular momenta.

$$|\, j_1, j_2 j_3 (J_{23}) : JM \rangle = \sum_{J_{12}} \langle j_1 j_2 (J_{12}); j_3 : JM \,|\, j_1, j_2 j_3 (J_{23}) : JM \rangle \times |\, j_1 j_2 (J_{12}), j_3 : JM \rangle$$

$$= \sum_{J_{12}} S^J_{J_{12}, J_{23}} (j_1 j_2 j_3) \,|\, j_1 j_2 (J_{12}), j_3 : JM \rangle. \qquad (5.97^1)$$

Applying the operators $\hat{J}_+$ and $\hat{J}_-$ to both sides of Eq. (5.97[1]), we get,

$$\hat{J}_\pm \,|\, j_1, j_2 j_3 (J_{23}) : JM > = \sum_{J_{12}} S^J_{J_{12}, J_{23}} (j_1 j_2 j_3) \hat{J}_\pm \,|\, j_1 j_2 (J_{12}), j_3 : JM >,$$

since $S^J_{J_{12}, J_{23}} (j_1 j_2 j_3)$ is only a number. From this, it follows that

$$|\, j_1, j_2 j_3 (J_{23}) : JM \pm 1 > = \sum_{J_{12}} S^J_{J_{12}, J_{23}} (j_1 j_2 j_3) \,|\, j_1 j_2 (J_{12}), j_3 : JM \pm 1 > \qquad (5.98)$$

From Eqs. (5.97[1]) and (5.98), it follows that

$$\langle j_1 j_2 (J_{12}), j_3 : JM \,|\, j_1, j_2 j_3 (J_{23}) : JM \rangle$$

is independent of M. We will denote this matrix element either by

$$U(j_1 j_2 J j_3 : J_{12} J_{23})$$

or by $$\{(2J_{12}+1)(2J_{23}+1)\}^{1/2} W(j_1 j_2 J j_3 : J_{12} J_{23}).$$

Thus,

$$< j_1 j_2(J_{12}), j_3 : JM \mid j_1, j_2 j_3(J_{23}) : JM > \equiv S^J_{J_{12} J_{23}}(j_1 j_2 j_3)$$

$$= \{(2J_{12}+1)(2J_{23}+1)\}^{1/2} W(j_1 j_2 J j_3 : J_{12} J_{23}) \tag{5.99a}$$

$$= U(j_1 j_2 J\, j_3 : J_{12} J_{23}). \tag{5.99b}$$

Here, $W(\ )$ is called the *Racah Coefficient*[14], while $U(\ )$ is known as the *normalized Racah Coefficient.*

Eq. (5.97[1]) now reads,

$$\mid j_1, j_2 j_3(J_{23}) : JM > = \sum_{J_{12}} U(j_1 J j_2 j_3 : J_{12} J_{23}) \mid j_1 j_2(J_{12}), j_3 : JM > . \tag{5.97[2]}$$

We will see below [Eq. (5.102)] that the Racah coefficients are real. Therefore, the unitarity of the matrix $S^J(j_1 j_2 j_3)$ is expressed by the relations,

$$\sum_{J_{23}} S^J_{J_{12}, J_{23}}(j_1 j_2 j_3)\, S^J_{J_{12}', J_{23}}(j_1 j_2 j_3) = \sum_{J_{23}} U(j_1, j_2 J j_3 : J_{12} J_{23})$$

$$U(j_1 j_2 J j_3 : J'_{12} J_{23}) = \delta_{J_{12} J'_{12}} \tag{5.100a}$$

and

$$\sum_{J_{12}} S^J_{J_{12}, J_{23}}(j_1, j_2 j_3)\, S^J_{J_{12}, J_{23}}(j_1, j_2 j_3) = \sum_{J_{12}} U(j_1 j_2 J j_3 : J_{12} J_{23})$$

$$U(j_1 j_2 J j_3 : J_{12} J'_{23}) = \delta_{J_{23} J'_{23}} \tag{5.100b}$$

Using Eq. (5.100a), we can invert the relationship (5.97[2]) to obtain,

$$\mid j_1 j_2(J_{12}), j_3 : JM > = \sum_{J_{23}} U(j_1 j_2 J j_3 : J_{12} J_{23}) \mid j_1, j_2 j_3(J_{23}) : JM > . \tag{5.97[3]}$$

Now, multiplying both sides of Eq. (5.95[1]) by $C^{j_1 j_2 J_{12}}_{m_1 m_2 M_{12}} \times C^{J_{12} j_3 J}_{M_{12} m_3 M}$ and summing over J_{12} and J, we get, using Eq. (5.84b),

$$\mid j_1 j_2 j_3 m_1 m_2 m_3 > = \sum_{J_{12}, J} C^{j_1 j_2 J_{12}}_{m_1 m_2 M_{12}} C^{J_{12} j_3 J}_{M_{12} m_3 M} \times \mid j_1 j_2(J_{12}), j_3 : JM > . \tag{5.101}$$

Substituting from Eq. (5.101) in Eq. (5.95[2]), we have,

$$\mid j_1, j_2 j_3(J_{23}) : JM > = \sum_{\substack{m_1 m_2 m_3 \\ J_{12}}} C^{j_2 j_3 J_{23}}_{m_2 m_3 M_{23}} C^{j_1 J_{23} J}_{m_1 M_{23} M} \times C^{j_1 j_2 J_{12}}_{m_1 m_2 M_{12}} C^{J_{12} j_3 J}_{M_{12} m_3 M} \mid j_1 j_2(J_{12}),\ j_3 : JM > .$$

(5.97[4])

where, the summation over J has been omitted because of the occurrence of J on the left hand side. Comparing Eq. (5.97[4]) with Eq. (5.97[2]), we see that,

14. Recah, G. [Footnote 11] was the first to introduce these.

$$U(j_1j_2Jj_3;J_{12}J_{23}) = \sum_{m_1m_2m_3} C^{j_1j_2J_{12}}_{m_1m_2M_{12}} C^{J_{12}j_3J}_{M_{12}m_3M} C^{j_2j_3J_{23}}_{m_2m_3M_{23}} C^{j_1J_{23}J}_{m_1M_{23}M} \tag{5.102}$$

From the reality of the C-coefficients, it follows that the *Racah coefficients are real.* It also shows that $U(abcd: ef)$ vanishes unless the triangular conditions $\Delta(abe), \Delta(edc), \Delta(bdf), \Delta(afc)$, are satisfied, where

$$\Delta(lmn) = \left[\frac{(l+m-n)!(l-m+n)!(-l+m+n)!}{(l+m+n+1)!}\right]^{1/2}, \tag{5.103}$$

Racah (Footnote 11) has given the following explicit expression for the W-coefficient :

$$W(abcd:ef) = \Delta(abe)\Delta(edc)\Delta(bdf)\Delta(afc)$$

$$\times \sum_{\kappa}(-1)^{\kappa}\frac{(a+b+c+d+1-\kappa)!}{\kappa!(a+b-e-\kappa)!(c+d-e-\kappa)!(a+c-f-\kappa)!}$$

$$\times\frac{1}{(b+d-f-\kappa)!(\kappa+e+f-a-d)!(\kappa+e+f-b-c)!} \tag{5.104}$$

Here, κ takes all values consistent with the factorial notation : That is,

$$\kappa_{\text{minimum}} = \text{ larger of } (a+d-e-f) \text{ and } (b+c-e-f);$$

$$\kappa_{\text{maximum}} = \text{ smallest of } (a+b-e), (c+d-e), (a+c-f)$$

$$\text{and } (b+d-f).$$

A procedure similar to the one that led to Eq. (5.102) may be used, along with the properties of the C-coefficients, to derive the following relationships:

$$\sum_{J_{12}} U(j_1j_2Jj_3;J_{12}J_{23}) C^{j_1j_2J_{12}}_{m_1m_2M_{12}} C^{J_{12}j_3J}_{M_{12}m_3M} = C^{j_2j_3J_{23}}_{m_2m_3M_{23}} C^{j_1J_{23}J}_{m_1M_{23}M} \tag{5.105a}$$

$$U(j_1j_2Jj_3;J_{12}J_{23}) C^{J_{12}j_3J}_{M_{12}m_3M} = \sum_{m_1m_2} C^{j_1j_2J_{12}}_{m_1m_2M_{12}} C^{j_2j_3J_{23}}_{m_2m_3M_{23}} C^{j_1J_{23}J}_{m_1M_{23}M} \tag{5.105b}$$

Definition (5.87) may be used to derive from (5.105a), the relationship,

$$\sum_{k} D_{l_1'l_1k} D_{l_2'l_2k} W(l_1'l_1l_2'l_2;kL) = (-1)^{l_1+l_1'-L} D_{l_1l_2L} D_{l_1'l_2'L}. \tag{5.106}$$

Properties of the Racah Coefficients

Symmetry: The symmetry properties follow from the explicit expression (5.104) for the Racah coefficient.

(i) $(-1)^{-e-f}W(abcd:ef)$ is invariant under all transformations which maintain the same triangles (see Fig. 5.7).

Thus,

$$(-1)^{-e-f}W(abcd:ef) = (-1)^{-b-c}W(aefd:bc)$$

$$= (-1)^{-a-d}W(ebcf:ad). \tag{5.107a}$$

(ii) $W(abcd:ef)$ is invariant under an even permutation of the arguments.

$$W(abcd:ef) = W(badc:ef)$$
$$= W(dcba:ef)$$
$$= W(acbd:fe) \qquad (5.107b)$$

A few others could be obtained by repeated application of these.

(iii) A third class of symmetry relations could be better expressed in terms of the 6*j*-symbols, defined by,

$$\begin{Bmatrix} j_1 j_2 j_3 \\ l_1 l_2 l_3 \end{Bmatrix} = (-1)^{j_1+j_2+l_1+l_2} W(j_1 j_2 l_2 l_1 : j_3 l_3). \qquad (5.108)$$

The symmetry relations (5.107a, b) correspond to the invariance of the 6*j*-symbols under (a) interchange of any two columns, (b) inversion of any two columns.

In order to express the third type of symmetry concisely, we define,

$$A = j_1 + l_1;\ \alpha = j_1 - l_1;\ A_{\pm\alpha} = \frac{1}{2}(A \pm \alpha)$$

$$B = j_2 + l_2;\ \beta = j_2 - l_2;\ B_{\pm\alpha} = \frac{1}{2}(B \pm \alpha) \qquad (5.109)$$

$$C = j_3 + l_3;\ \gamma = j_3 - l_3;\ C_{\pm\alpha} = \frac{1}{2}(C \pm \alpha),\ \text{etc.}$$

Then, $\begin{Bmatrix} j_1 j_2 j_3 \\ l_1 l_2 l_3 \end{Bmatrix} \equiv \begin{Bmatrix} A_{+\alpha} B_{+\beta} C_{+\gamma} \\ A_{-\alpha} B_{-\beta} C_{-\gamma} \end{Bmatrix}$, is invariant under the columnwise permutations of the alphabets A, B, C. Thus,

$$\begin{Bmatrix} A_{+\alpha} B_{+\beta} C_{+\gamma} \\ A_{-\alpha} B_{-\beta} C_{-\gamma} \end{Bmatrix} = \begin{Bmatrix} B_{+\alpha} A_{+\beta} C_{+\gamma} \\ B_{-\alpha} A_{-\beta} C_{-\gamma} \end{Bmatrix} = \begin{Bmatrix} B_{+\alpha} C_{+\beta} A_{+\gamma} \\ B_{-\alpha} C_{-\beta} A_{-\gamma} \end{Bmatrix}. \qquad (5.107c)$$

Orthogonality and Sum Rules

The orthogonality relations follow from the unitarity of the matrix $S^{J(j_1 j_2 j_3)}_{J_{12} J_{23}}$, and are given by Eqs. (5.100a, b). The sum rules could be derived with the help of the orthogonality relations :

$$\sum_f (-1)^f U(abcd:ef) U(adcb:gf) = (-1)^s U(abdc:eg), \qquad (5.110a)$$

where,

$$s = e + g - a - b - c - d.$$

$$\sum_f 2U(abcd:ef) U(abcd:gf) = \delta_{eg} \pm (-1)^{e+g-b-d} U(a\bar{b}dc:eg), \qquad (5.110b)$$

where the upper sign is for odd f and the lower sign for even f.

$$\sum_{\lambda}(2\lambda+1)\,W(a'\lambda be:ae')\,W(c\lambda de':c'e)\,W(a'\lambda fc:ac') = W(abcd:ef)\,W(a'bc'd:e'f), \tag{5.110c}$$

$$\sum_{e}(-1)^{e}(2e+1)\,W(abab:ef) = \delta_{f,0}(-1)^{a+b}\{(2a+1)(2b+1)\}^{1/2} \tag{5.110d}$$

Recursion Relations

These also could be expressed more conveniently in terms of the 6*j*-symbols.

$$(-1)^{x+y-j_1-j_2-j_3}\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix}\begin{Bmatrix} j_2 & l_3 & l_1 \\ \frac{1}{2} & l_1+x & l_3+y \end{Bmatrix}$$

$$= 2l_2\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1+x & l_2+\frac{1}{2} & l_3+y \end{Bmatrix}\begin{Bmatrix} j_1 & l_3 & l_2 \\ \frac{1}{2} & l_2-\frac{1}{2} & l_3+y \end{Bmatrix}\begin{Bmatrix} j_3 & l_1 & l_2 \\ \frac{1}{2} & l_2-\frac{1}{2} & l_1+x \end{Bmatrix}$$

$$-(2l_2+1)\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1+x & l_2+\frac{1}{2} & l_3+y \end{Bmatrix}\begin{Bmatrix} j_1 & l_3 & l_2 \\ \frac{1}{2} & l_2+\frac{1}{2} & l_3+y \end{Bmatrix}\begin{Bmatrix} j_3 & l_1 & l_2 \\ \frac{1}{2} & l_2+\frac{1}{2} & l_1+x \end{Bmatrix}$$

where, $x = \pm\frac{1}{2}$; $y = \pm\frac{1}{2}$. (5.111)

Problem 5.16: Deduce the following:

$$w(abcd:0f) = (-1)^{b+c-f}\{(2b+1)(2c+1)\}^{-1/2}\delta_{ab}\delta_{cd}.$$

5.5C The 9*j*-Symbols

The Clebsch-Gordon coefficients or the 3*j*-symbols give us the relationship between two sets of basis vectors corresponding to the coupled and the uncoupled representations of a system consisting of two individual angular momenta. The three *j*'s refer to the two individual angular momenta and their resultant. The Racah coefficient or the 6*j*-symbols, on the other hand, describe the relationship between two sets of basis vectors both of which correspond to coupled representations of a system consisting of three individual angular momenta. The six *j*'s are the three individual *j*'s, their resultant and the two *intermediate* angular momenta. For, the addition of three angular momenta is accomplished by adding two of them to obtain an intermediate angular momentum and then adding the third to this intermediate one to obtain the final resultant value. Since the different representations (bases) correspond to different intermediate angular momenta and since the 6*j*-symbols relate to two different representations, they (the 6*j*-symbols) would involve two intermediate angular momenta.

Extending the above arguments, we see that the addition of n angular momenta is accomplished through the introduction of $(n-2)$ intermediate angular momenta. This implies that the transformation coefficients (the elements of the transforming matrix) relating any two different coupling schemes would involve a total of $\{(n+1)+2(n-2)\} = 3(n-1), j$'s. These coefficients are, therefore, called $3(n-1)$ *j-symbols*. Thus, for $n = 4$, we have the $9j$-symbols, for $n=5$, the $12j$-symbols, and so on. In view of the importance of the $9j$-symbols in certain branches of physics governed by quantum mechanics, we will discuss them here; but no higher j-symbols will be discussed.

In analogy with Eq. (5.99a), the $9j$-symbol, or *X-coefficient*, is defined by,

$$< j_1 j_2(J_{12}),\ j_3 j_4(J_{34}): JM \mid j_1 j_3(J_{13}),\ j_2 j_4(J_{24}): JM >$$

$$\equiv X^{J(j_1 j_2 j_3 j_4)}_{J_{12} J_{34}: J_{13} J_{24}} = \{[J_{12}]\,[J_{34}]\,[J_{13}]\,[J_{24}]\}^{1/2} \begin{Bmatrix} j_1 & j_2 & J_{12} \\ j_3 & j_4 & J_{34} \\ J_{13} & J_{24} & J \end{Bmatrix}, \tag{5.112}$$

where

$$[j] \equiv (2j+1). \tag{5.113}$$

Eq. (5.112) implies

$$\mid j_1 j_3(J_{13}), j_2 j_4(J_{24}): JM > =$$

$$\sum_{J_{12}, J_{34}} X^{J(j_1 j_2 j_3 j_4)}_{J_{12} J_{34}: J_{13} J_{24}} \times \mid j_1 j_2(J_{12}), j_3 j_4(J_{34}): JM >, \tag{5.114[1]}$$

and, because of the unitarity of the matrix $X^J(j_1 j_2 j_3 j_4)$,

$$\mid j_1 j_2(J_{12}), j_3 j_4(J_{34}): JM >$$

$$= \sum_{J_{13} J_{24}} X^{J(j_1 j_2 j_3 j_4)}_{J_{13} J_{24}: J_{12} J_{34}} \mid j_1 j_3(J_{13}), j_2 j_4(J_{24}): JM > \tag{5.115}$$

But, using Eqs. (5.97[2]) and (5.97[3]), we have,

$$\mid j_1 j_3(J_{13}), j_2 j_4(J_{24}): JM >$$

$$= \sum_{\lambda} U(j_1 j_3 J J_{24}: J_{13} \lambda) \mid j_1, j_3 J_{24}(\lambda): JM >$$

$$= \sum_{\lambda, J_{34}} U(j_1 j_3 J J_{24}: J_{13} \lambda)\,(-1)^{j_2 + j_4 - J_{24}}$$

$$\times U(j_3 j_4 \lambda j_2: J_{34} J_{24}) \mid j_1, J_{34} j_2(\lambda): JM >$$

$$= \sum_{\lambda, J_{34}, J_{12}} (-1)^{j_2 + j_4 - J_{24}} U(j_1 j_3 J J_{24}: J_{13} \lambda)$$

$$\times U(j_3 j_4 \lambda j_2: J_{34} J_{24})\,(-1)^{\lambda - j_2 - J_{34}} U(j_1 j_2 J J_{34}: J_{12} \lambda)$$

$$\times \mid j_1 j_2(J_{12}), j_3 j_4(J_{34}): JM >$$

$$= \sum_{J_{12},J_{34}} \{[J_{12}]\,[J_{34}]\,[J_{13}]\,[J_{24}]\}^{1/2} \sum_{\lambda}(2\lambda+1)$$

$$\times W(j_1 j_2 J J_{34}; J_{12}\lambda) W(j_1 j_3 J J_{24}; J_{13}\lambda) \qquad (5.114^2)$$

$$\times W(j_2 J_{24} J_{34} j_3; j_4 \lambda) \mid j_1 j_2 (J_{12}), j_3 j_4 (J_{34}); JM >,$$

where, Eqs. (5.99a, b) and (5.107b) have been used.

From Eqs. (5.114[1-2]), we get,

$$\begin{Bmatrix} j_1 & j_2 & J_{12} \\ j_3 & j_4 & J_{34} \\ J_{13} & J_{24} & J \end{Bmatrix} = \sum_{\lambda}(2\lambda+1) W(j_1 j_2 J J_{34}; J_{12}\lambda) W(j_1 j_3 J J_{24}; J_{13}\lambda) \times W(j_2 J_{24} J_{34} j_3; j_4 \lambda)$$

(5.116a)

In terms of the 6*j*-symbols [Eq. (5.108)], we have,

$$\begin{Bmatrix} j_1 & j_2 & J_{12} \\ j_3 & j_4 & J_{34} \\ J_{13} & J_{24} & J \end{Bmatrix} = \sum_{\lambda}(-1)^{2\lambda}(2\lambda+1) \begin{Bmatrix} j_1 & j_3 & J_{13} \\ J_{24} & J & \lambda \end{Bmatrix} \begin{Bmatrix} j_2 & j_4 & J_{24} \\ j_3 & \lambda & J_{34} \end{Bmatrix}$$

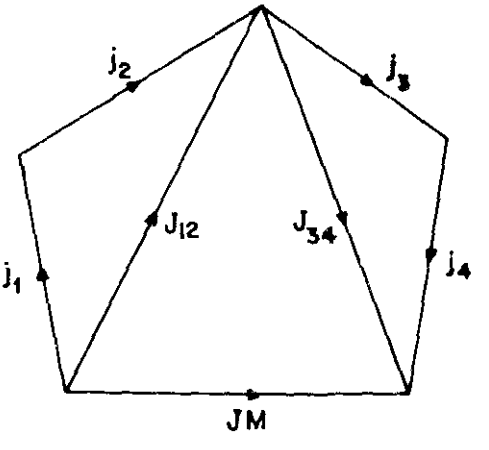
$$\times \begin{Bmatrix} J_{12} & J_{34} & J \\ \lambda & j_1 & j_2 \end{Bmatrix}. \qquad (5.116b)$$

Similarly, it can be shown that,

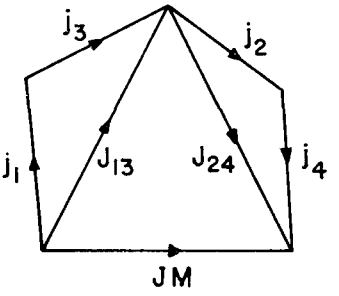
$$\begin{Bmatrix} j_1 & j_2 & J_{12} \\ j_3 & j_4 & J_{34} \\ J_{13} & J_{24} & J \end{Bmatrix} = \{[J_{12}]\,[J_{34}]\,[J_{13}]\,[J_{24}]\}^{-1/2} \times \sum_{\text{All } m\text{'s}} C^{j_1 j_2 J_{12}}_{m_1 m_2 M_{12}} C^{j_3 j_4 J_{34}}_{m_3 m_4 M_{34}} \qquad (5.117)$$

$$\times C^{J_{13} J_{24} J}_{M_{13} M_{24} M} C^{j_1 j_3 J_{13}}_{m_1 m_3 M_{13}} C^{j_2 j_4 J_{24}}_{m_2 m_4 M_{24}} C^{J_{12} J_{34} J}_{M_{12} M_{34} M}$$

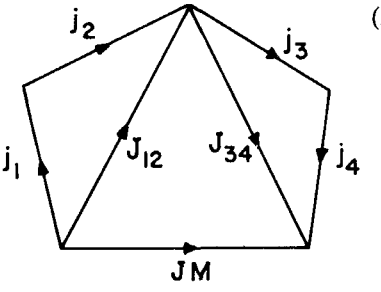
Now, the wavefunction $\mid j_1 j_2 (J_{12}), j_3 j_4 (J_{34}); JM >$ could be schematically represented as,



Equation (5.114[1]) would be then written as,

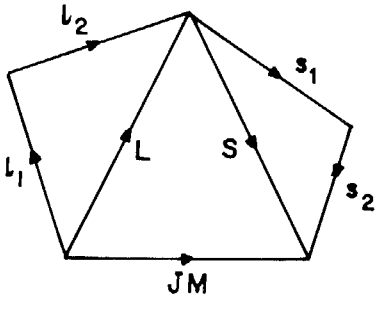


$$= \sum_{J_{12},J_{34}} X^{J(j_1j_2j_3j_4)}_{J_{12}J_{34}J_{13}J_{24}}$$

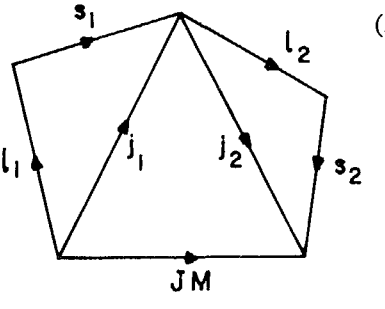


(5.118)

In particular, the relationship between the wavefunctions in the "*LS*-coupling scheme", and the "*jj*-coupling scheme" for a system consisting of two particles (say, nucleons) with individual orbital and spin angular momenta given respectively, by l_1, s_1 and $l_2 s_2$, can be expressed as,



$$= \sum_{j_1,j_2} X^{J(l_1s_1l_2s_2)}_{j_1j_2;LS}$$



(5.119)

Writing the 9 j-symbol as $\begin{Bmatrix} a & b & e \\ c & d & f \\ g & h & k \end{Bmatrix}$,

We can summarise the important properties as follows. These properties can be derived from the properties of angular momenta, the orthogonality and completeness of the basis vectors and similar general properties.

Selection Rules

The coefficient vanishes unless each row and column add up to an integral number and also satisfies the triangular conditions $\Delta(abe), \Delta(acg)$, etc. It follows from this that if any two rows (columns) are identical, then the coefficient vanishes unless the sum of the third row (column) is even.

Symmetry

(i) The coefficient is invariant under the interchange of rows and columns (transposition).

(ii) An odd permutation of the rows or columns multiples the coefficient by $(-1)^{\epsilon}$ where ϵ is the sum of all the $9j$'s.

Orthogonality and Sum Rules:

$$\sum_{e,f} [e]\,[f]\,[g]\,[h] \begin{Bmatrix} a & b & e \\ c & d & f \\ g & h & k \end{Bmatrix} \begin{Bmatrix} a & b & e \\ c & d & f \\ g' & h' & k \end{Bmatrix} = \delta_{gg'}\delta_{hh'} \tag{5.120a}$$

$$\sum_{g,h} [e]\,[f]\,[g]\,[h] \begin{Bmatrix} a & b & e \\ c & d & f \\ g & h & k \end{Bmatrix} \begin{Bmatrix} a & b & e' \\ c & d & f' \\ g & h & k \end{Bmatrix} = \delta_{ee'}\delta_{gg'} \tag{5.120b}$$

$$\sum_{g,h} (-1)^{2b+h+f'-f} [g]\,[h] \begin{Bmatrix} a & b & e \\ c & d & f \\ g & h & k \end{Bmatrix} \begin{Bmatrix} a & c & g \\ d & b & h \\ e' & f' & k \end{Bmatrix} = \begin{Bmatrix} a & b & e \\ d & c & f \\ e' & f' & k \end{Bmatrix} \tag{5.121a}$$

$$\sum_{e} (2e+1) \begin{Bmatrix} a & b & e \\ c & d & f \\ g & h & k \end{Bmatrix} W(abkf:e\lambda) = W(bhfc:d\lambda) W(khac:g\lambda). \tag{5.121b}$$

$$\sum_{d,e,f} [d]\,[e]\,[f] \begin{Bmatrix} a & b & c \\ a & b & c \\ d & e & f \end{Bmatrix} = \frac{1}{4}\left\{1 + \frac{(-1)^{2a}}{2a+1} + \frac{(-1)^{2b}}{2b+1} + \frac{(-1)^{2c}}{2c+1}\right\}. \tag{5.121c}$$

In connection with the LS- and jj-coupling schemes, the following relationship would be useful.

$$\begin{Bmatrix} s & s & 1 \\ l_1 & l_2 & L \\ j_1 & j_2 & L \end{Bmatrix} = \frac{l_1(l_1+1)+j_2(j_2+1)-l_2(l_2+1)-j_1(j_1+1)}{2\{s(s+1)(2s+1)L(L+1)(2L+1)\}^{1/2}} \begin{Bmatrix} s & s & 0 \\ l_1 & l_2 & L \\ j_1 & j_2 & L \end{Bmatrix} \tag{5.122}$$

When one of the arguments (j's) of the $9j$-symbol is zero, the latter reduces to a Racah coefficient. For example,

$$\begin{Bmatrix} a & b & e \\ c & d & f \\ g & h & 0 \end{Bmatrix} = \delta_{ef}\delta_{gh}\{(2e+1)(2g+1)\}^{-1/2} W(ebcg:ad). \tag{5.123}$$

Problem 5.17:

(a) Deduce Eq. (5.121c) from Eq. (5.117)

(b) Derive Eq. (5.123) using Eq. (5.116a).

5.6 ANGULAR MOMENTUM AND ROTATIONS

A rotation is specified by an *axis*, usually denoted by a unit vector **n**, about which, and an *angle* ϕ through which, the rotation is made. The positive sense of rotation is defined by the right-handed-screw-rule[15]. Thus, $(\mathbf{n}, \phi)$ represents a positive rotation about the axis **n** through an angle ϕ, whereas $(\mathbf{n}, -\phi)$ is a negative rotation about **n** through the angle ϕ.

Now, a rotation could be of the following two types:

(i) Rotation of the physical system, with the co-ordinate system fixed in space.

(ii) Rotation of the co-ordinate system with the physical system fixed in space.

Unless otherwise specified, in this section, we mean by 'rotation', a rotation of the physical system. Of course, since only the *relative* orientations of the physical system and the co-ordinate system are of relevance to the description of physical systems, a rotation $(\mathbf{n}, \phi)$ of the physical system is equivalent to the rotation $(\mathbf{n}, -\phi)$ of the co-ordinate system.

Now, rotations are linear transformations in the physical space. Therefore, it is possible to define a linear operator $\hat{R}_{\mathbf{n}}(\phi)$ corresponding to a rotation $(\mathbf{n}, \phi)$ of the physical system. The rotation $(\mathbf{n}, -\phi)$ is, then, represented by the operator,

---

15. A right-handed screw advances in the direction of **n** when twisted in the sense of the positive rotation.

$$\hat{R}_n(-\phi) = \hat{R}_n^{-1}(\phi). \tag{5.124}$$

In order to illustrate the effect of a rotation on the wavefunction of the system let us consider specifically a rotation about the z-axis through an angle ϕ. We will denote the corresponding rotation operator by $\hat{R}_z(\phi)$. Let $\psi(\mathbf{r}) \equiv \psi(x, y, z)$, represent the value of the wavefunction at the (space) point P (Fig 5.8) before the rotation; and $\psi'(\mathbf{r})$ the value at the *same* point after the rotation. Then, by definition, we have,

$$\hat{R}_z(\phi)\,[\psi(\mathbf{r})] = \psi'(\mathbf{r}). \tag{5.125^1}$$

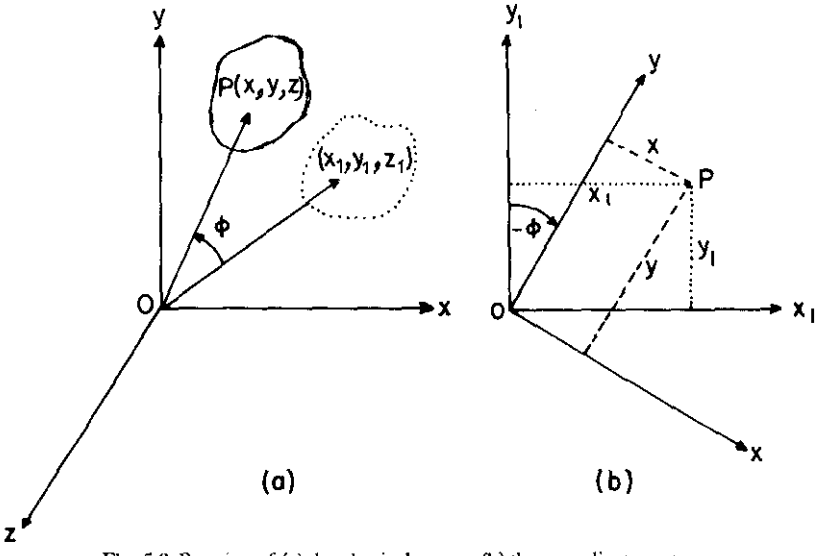


Fig. 5.8. Rotation of (a) the physical system (b) the co-ordinate system.

But the value of the wavefunction at the point P cannot depend on the position of P in space (that is, on its co-ordinates with respect to the co-ordinate system fixed in space), but should depend only on the point's relative position in the physical system (that is, on the co-ordinates of P referred to a co-ordinate frame fixed to the physical system). Therefore, the value of $\psi'(\mathbf{r})$ is equal to the value of ψ at the point $\mathbf{r}_1$, where $\mathbf{r}_1$ was the position of P before the rotation. That is,

$$\mathbf{r}_1 = \hat{R}_z(-\phi)\,\{\mathbf{r}\} = \hat{R}_z^{-1}(\phi)\,\{\mathbf{r}\} = \hat{K}_z(\phi)\,\{\mathbf{r}\}, \tag{5.126}$$

where $\hat{K}_z(\phi)$ represents a rotation of the co-ordinate system through an angle ϕ about the z-axis.

Thus,

$$\hat{R}_z(\phi)\,\{\psi(\mathbf{r})\} = \psi'(\mathbf{r}) = \psi(\mathbf{r}_1) = \psi[\hat{R}_z^{-1}(\phi)\,\{\mathbf{r}\}] \tag{5.125^2}$$

From Fig. (5.8b), we have,

$$x_1 = x\cos\phi + y\sin\phi,$$

$$y_1 = -x\sin\phi + y\cos\phi,$$

$$z_1 = z. \tag{5.126a}$$

Infinitesimal Rotations

Replacing finite rotation ϕ by an infinitesimal rotation $\delta\phi$ in (5.126a), we have,

$$x_1 = x\cos\delta\phi + y\sin\delta\phi \approx x + y\delta\phi,$$

$$y_1 = -x\sin\delta\phi + y\cos\delta\phi \approx -x\delta\phi + y,$$

$$z_1 = z. \tag{1.126b}$$

Substituting in Eq. (5.125^2) from (5.126b), we get,

$$\hat{R}_z(\delta\phi)\{\psi(x,y,z)\} = \psi(x + y\delta\phi, y - x\delta\phi, z)$$

$$\approx \psi(x,y,z) + \delta\phi\left(y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right)\psi$$

$$= \left(1 - \frac{i}{\hbar}\delta\phi\hat{L}_z\right)\psi(x,y,z), \tag{5.127^1}$$

where Taylor series expansion, and the definition (5.40) for the orbital angular momentum operator, have been used.

Alternatively, writing $\psi(\mathbf{r}) \equiv \psi(r,\theta,\phi)$, where (r,θ,ϕ) are the spherical co-ordinates, we have, from Eq. (5.125^2),

$$\hat{R}_z(\delta\phi)\{\psi(r,\theta,\phi)\} = \psi(r_1,\theta_1,\phi_1) = \psi(r,\theta,\phi - \delta\phi)$$

$$= \left(\hat{1} - \delta\phi\frac{\partial}{\partial\phi}\right)\psi(r,\theta,\phi), \tag{5.127^1}$$

which reduces to Eq. (5.127^1), when definition (5.42c) is used.

Thus, $$\hat{R}_z(\delta\phi) = \left(\hat{1} - \frac{i}{\hbar}\delta\phi\cdot\hat{L}_z\right). \tag{5.128^1}$$

Similarly, an infinitesimal rotation $\delta\theta$ about an axis $\mathbf{n}$, is represented by the operator,

$$\hat{R}_n(\delta\theta) = \hat{1} - \frac{i}{\hbar}\delta\theta(\mathbf{n}\cdot\hat{\mathbf{L}}). \tag{5.128^2}$$

In the derivation of Eq. (5.128^2), we have assumed that the wavefunction of the system is a function of the space co-ordinates only. If the system has also intrinsic angular momentum $\mathbf{S}$ in addition to $\mathbf{L}$, then, the total wavefunction would be a product of $\psi(\mathbf{r})$ and χ_s, where χ_s is the spin part of the wavefunction. We now *postulate* that χ_s transforms under the infinitesimal rotation $(\mathbf{n},\delta\theta)$ according to the formula (c.f. Eq. (5.127^1)),

$$\hat{R}_n(\delta\theta)\{\chi_s\} = \left[\hat{1} - \frac{i}{\hbar}\delta\theta(\mathbf{n}\cdot\hat{\mathbf{S}})\right]\chi_s. \tag{5.127^2}$$

Then, the total wavefunction $\psi(\mathbf{r})\chi_s$ would transform according to,

$$\hat{R}_{\mathbf{n}}(\delta\theta)\{\psi(\mathbf{r})\chi_s\} = \left\{\hat{1} - \frac{i}{\hbar}\delta\theta(\mathbf{n}\cdot\hat{\mathbf{L}})\right\}\left\{\hat{1} - \frac{i}{\hbar}\delta\theta(\mathbf{n}\cdot\hat{\mathbf{S}})\right\}\psi(\mathbf{r})\chi_s$$

$$\approx \left[\hat{1} - \frac{i}{\hbar}\delta\theta(\mathbf{n}\cdot\hat{\mathbf{J}})\right]\psi(\mathbf{r})\chi_s; \qquad (5.127^3)$$

where $\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}$, is the operator corresponding to the total angular momentum. Thus, for a system with a total angular momentum $\hat{\mathbf{J}}$ the infinitesimal rotation operator is given by

$$\hat{R}_{\mathbf{n}}(\delta\theta) \equiv \hat{1} - \frac{i}{\hbar}(\mathbf{n}\cdot\hat{\mathbf{J}})\delta\theta. \qquad (5.128^3)$$

In fact, this equation could be regarded as a definition of the angular momentum operator $\hat{\mathbf{J}}$, since it leads to the basic commutation relations (5.7a) for the components of $\hat{\mathbf{J}}$. This could be established as follows:

Since $\hat{\mathbf{J}}$ in Eq. (5.128^3) is Hermitian, $\hat{R}_{\mathbf{n}}(\delta\theta)$ is Unitary [see Eq. (2.63)] so that an infinitesimal rotation represents a unitary transformation. A vector operator $\hat{\mathbf{V}}$, therefore, transforms under the infinitesimal rotation $(\mathbf{n}, \delta\theta)$ according to[16]

$$\hat{\mathbf{V}} \to \hat{\mathbf{V}}' = \hat{R}_{\mathbf{n}}(\delta\theta)\hat{\mathbf{V}}\hat{R}^{\dagger}_{\mathbf{n}}(\delta\theta)$$

$$= \hat{\mathbf{V}} - \frac{i}{\hbar}\delta\theta[(\mathbf{n}\cdot\hat{\mathbf{J}}), \hat{\mathbf{V}}], \qquad (5.129)$$

since $$\hat{R}^{\dagger}_{\mathbf{n}}(\delta\theta) = \hat{R}^{-1}_{\mathbf{n}}(\delta\theta) = \hat{R}_{\mathbf{n}}(-\delta\theta). \qquad (5.130)$$

Thus, the change in $\hat{\mathbf{V}}$ is given by

$$\delta\hat{\mathbf{V}} \equiv \hat{\mathbf{V}}' - \hat{\mathbf{V}} = -(i/\hbar)\delta\theta[(\mathbf{n}\cdot\hat{\mathbf{J}}), \hat{\mathbf{V}}]. \qquad (5.131)$$

The change in the component of $\hat{\mathbf{V}}$ along the unit vector $\mathbf{u}$ is, therefore, given by,

$$\delta\hat{V}_{\mathbf{u}} \equiv \mathbf{u}\cdot\delta\hat{\mathbf{V}} = -i/\hbar)\delta\theta[(\mathbf{n}\cdot\hat{\mathbf{J}}), (\mathbf{u}\cdot\hat{\mathbf{V}})]. \qquad (5.132^1)$$

But $$\mathbf{u}\cdot\delta\hat{\mathbf{V}} = \delta\mathbf{u}\cdot\hat{\mathbf{V}}. \qquad (5.132^2)$$

$\delta\mathbf{u}$ being the change (as viewed from the space frame) in the unit vector $\mathbf{u}$ (fixed with respect to the physical system) resulting from the rotation $(\mathbf{u}, \delta\theta)$ of the physical system. That is [see Fig. (5.9)],

$$\delta\mathbf{u} = \delta\theta(\mathbf{n}\times\mathbf{u}). \qquad (5.133)$$

---

16. Note that, in terms of the operator $\hat{R}_{\mathbf{n}}$ representing rotations of the reference frame [cf. Eq. (2.126)],

$$\hat{\mathbf{V}}' = \hat{R}^{\dagger}_{\mathbf{n}}(\delta\theta)\hat{\mathbf{V}}\hat{R}_{\mathbf{n}}(\delta\theta).$$

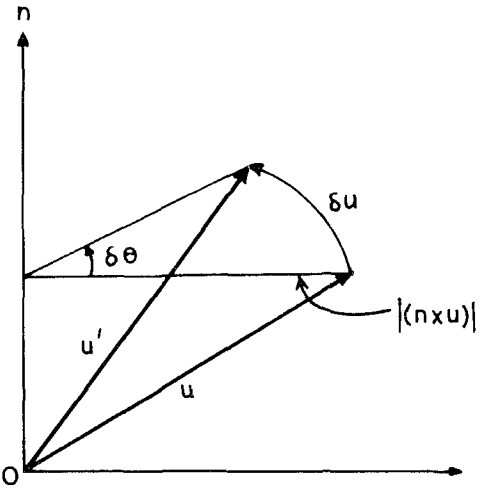


Fig. 5.9. Effect of rotation $\delta\theta$ about $\mathbf{n}$ on a unit vector $\mathbf{u}$.

Hence,

$$\delta\hat{\mathbf{V}}_u = \delta\theta(\mathbf{n}\times\mathbf{u})\cdot\hat{\mathbf{V}} \qquad (5.132^3)$$

From Eqs. (5.132^1) and (5.132^3), we get,

$$[(\mathbf{n}\cdot\hat{\mathbf{J}}),(\mathbf{u}\cdot\hat{\mathbf{V}})] = i\hbar(\mathbf{n}\times\mathbf{u})\cdot\hat{\mathbf{V}}. \qquad (5.134a)$$

Substituting $\mathbf{i}$ for $\mathbf{n}$, $\mathbf{j}$ for $\mathbf{u}$ and $\hat{\mathbf{J}}$ for $\hat{\mathbf{V}}$ in this equation, we get Eq. (5.7a).

It should be emphasised that only the operator corresponding to the *total* angular momentum of the system would satisfy Eq. (5.134a) for *all* the vector operators. For example, in the case of a system of N particles, the angular momenta of the individual particles as well as the sum of the angular momenta of a few of the particles, satisfy the relationship (5.7a) but not the relationship (5.134a). The operator corresponding to the total angular momentum of a system is, thus, the *generator* of infinitesimal rotations of the system.

A scalar, by definition, is invariant under rotations. Therefore, a scalar operator $\hat{S}$ transforms under the infinitesimal rotation $(\mathbf{n}, \delta\theta)$ according to

$$\hat{S} \to \hat{S}' = \hat{R}_{\mathbf{n}}(\delta\theta)\hat{S}\hat{R}_{\mathbf{n}}^{\dagger}(\delta\theta) = \hat{S},$$

or,

$$\hat{R}_{\mathbf{n}}\hat{S} = \hat{S}\hat{R}_{\mathbf{n}}.$$

That is,

$$[(\mathbf{n}\cdot\hat{\mathbf{J}}),\hat{S}] = \hat{0}. \qquad (5.134b)$$

Thus, a scalar operator commutes with the components of $\hat{\mathbf{J}}$.

Problem 5.18: If $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are two vector operators, show, using Eq. (5.134a), that

$$[(\mathbf{n}\cdot\hat{\mathbf{J}}),(\hat{\mathbf{A}}\cdot\hat{\mathbf{B}})] = \hat{0}.$$

Finite Rotations : Euler Angles

A finite rotation $(\mathbf{n}, \theta)$ could be regarded as a succession of a large number of infinitesimal rotations, so that,

$$\hat{R}_{\mathbf{n}}(\theta) = \underset{N\to\infty}{\text{Lt}} \left[\hat{1} - (i/\hbar)\frac{\theta}{N}(\mathbf{n}\cdot\hat{\mathbf{J}})\right]^N$$

$$= \sum_{p=0}^{\infty} \frac{(-1)^p}{p!}\{(i/\hbar)\theta(\mathbf{n}\cdot\hat{\mathbf{J}})\}^p$$

$$= \exp\left[-(i/\hbar)\theta\{(\mathbf{n}\cdot\hat{\mathbf{J}})\}\right]. \qquad (5.135)$$

Thus, a rotation through an angle α about the z-axis is represented by the operator $\hat{R}_z(\alpha) \equiv \exp\left[(-i/\hbar)\alpha\hat{J}_z\right]$. A general rotation, instead of being specified by the unit vector $\mathbf{n}$ (which requires two angles for its specification) and an angle, could be specified by three angles, usually known as the *Euler angles*, defined[17] as follows (see Fig. 5.10):

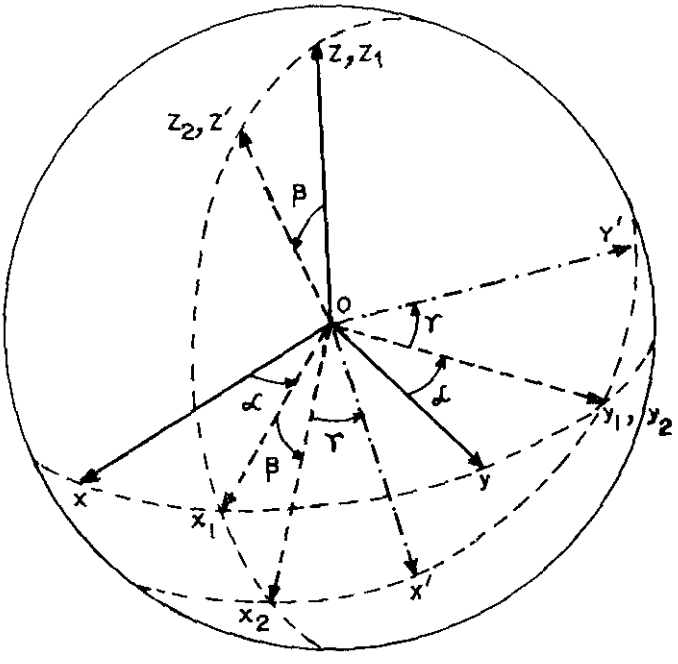


Fig. 5.10 The Euler angles.

(i) A rotation through α about the z_1-axis, in which

$$x \to x_1;\ y \to y_1;\ z \to z_1 = z.$$

17. Some authors adopt a definition in which the second rotation is about the x_1-axis instead of the y_1-axis. The definition adopted here is that of Ref. 2 listed at the end of the chapter.

(ii) A rotation through β about the new y- (that is, y_1-) axis.

$$x_1 \to x_2;\ y_1 \to y_2 = y_1;\ z_1 \to z_2.$$

(iii) A rotation through γ about the z_2–axis.

$$x_2 \to x';\ y_2 \to y';\ z_2 \to z' = z_2.$$

Thus,

$$\hat{R}_n(\theta) \equiv \hat{R}(\alpha\beta\gamma) = \hat{R}_{z_2}(\gamma)\hat{R}_{y_1}(\beta)\hat{R}_z(\alpha). \qquad (5.136^1)$$

Now, $\hat{R}_{y_1}(\beta)$ is the transform of $\hat{R}_y(\beta)$ under the previous rotation. That is [see Eq. (5.129)],

$$\hat{R}_{y_1}(\beta) = \hat{R}_z(\alpha)\hat{R}_y(\beta)\hat{R}_z^{-1}(\alpha). \qquad (5.137a)$$

Similarly,

$$\begin{aligned}\hat{R}_{z_2}(\gamma) &= \hat{R}_{y_1}(\beta)\hat{R}_{z_1}(\gamma)\hat{R}_{y_1}^{-1}(\beta)\\ &= \hat{R}_{y_1}(\beta)\hat{R}_z(\alpha)\hat{R}_z(\gamma)\hat{R}_z^{-1}(\alpha)\hat{R}_{y_1}^{-1}(\beta)\\ &= \hat{R}_z(\alpha)\hat{R}_y(\beta)\hat{R}_z(\gamma)\hat{R}_y^{-1}(\beta)\hat{R}_z^{-1}(\alpha)\end{aligned} \qquad (5.137b)$$

Substituting in (5.136^1) from (5.137a, b), we get[18],

$$\begin{aligned}\hat{R}(\alpha\beta\gamma) &= \hat{R}_z(\alpha)\hat{R}_y(\beta)\hat{R}_z(\gamma)\\ &= \exp[-(i/\hbar)\alpha\hat{\mathbf{J}}_z]\exp[-(i/\hbar)\beta\hat{\mathbf{J}}_y]\exp[-(i/\hbar)\gamma\hat{J}_z].\end{aligned} \qquad (5.136^2)$$

Problem 5.19: Show that

$$\exp[-(i/\hbar)\beta\hat{\mathbf{J}}_y] = \exp[(i/\hbar)(\pi/2)\hat{\mathbf{J}}_x]\exp[-(i/\hbar)\beta\hat{\mathbf{J}}_z]\exp[-(i/\hbar)(\pi/2)\hat{\mathbf{J}}_y].$$

The Rotation Matrix

In the representation spanned by the eigenvectors $|jm>$ of the angular momentum operator, the operator $\hat{R}$ would be represented by a square (unitary) matrix. We will now deduce the elements of this matrix.

Since $\hat{R}(\alpha\beta\gamma)$ contains only the angular momentum components $\hat{J}_y$ and $\hat{J}_z$, $\hat{R}(\alpha\beta\gamma)|jm>$ can differ from $|jm>$ only in the m-value and not in the value of j [see Eqs. (5.32b-e)]. That is,

$$\hat{R}(\alpha\beta\gamma)|jm> = \sum_{m'}|jm'><jm'|\hat{R}(\alpha\beta\gamma)|jm>. \qquad (5.138^1)$$

---

18. Note that, since $\hat{J}_x\hat{J}_y \neq \hat{J}_y\hat{J}_x$, $\exp[-(i/\hbar)\alpha\hat{J}_x]\exp[-(i/\hbar)\beta\hat{J}_y] \neq \exp[-(i/\hbar)(\alpha\hat{J}_x + \beta\hat{J}_y)]$

But, from Eq. (5.136[2]), we have,

$$< jm' \mid \hat{R}(\alpha\beta\gamma) \mid jm > = < jm' \mid \exp[-(i/\hbar)\alpha\hat{J}_z] \exp[-(i/\hbar)\beta\hat{J}_y] \exp[-(i/\hbar)\gamma\hat{J}_z] \mid jm >$$

$$= \exp[+(i/\hbar)\alpha\hat{J}_z]\psi_{jm'}, \exp[-(i/\hbar)\beta\hat{J}_y] \exp[-(i/\hbar)\gamma\hat{J}_z]\psi_{jm})$$

$$= \exp(-im'\alpha)d^j_{m'm}(\beta) \exp(-im\gamma), \tag{5.139}$$

where

$$d^j_{m'm}(\beta) = < jm' \mid \exp[-(i/\hbar)\beta\hat{J}_y] \mid jm > . \tag{5.140}$$

Now, if we denote the matrix representing $\hat{R}(\alpha\beta\gamma)$ in the $(2j + 1)$-dimensional representation spanned by the basis vectors $\{\mid jm >\}$ by $\mathcal{D}^j$, then, Eq. (5.138[1]) could be written as [see Eq. (2.118)],

$$\hat{R}(\alpha\beta\gamma) \mid jm > = \sum_{m'=-j}^{+j} \mathcal{D}j_{m'm}(\alpha\beta\gamma) \mid jm' >, \tag{5.138[2]}$$

or,

$$\hat{R}(\alpha\beta\gamma)\psi_{jm}(\mathbf{r}) = \psi_{jm}[\hat{R}^{-1}\mathbf{r}] = \sum_{m'} \mathcal{D}^j_{m'm}(\alpha\beta\gamma)\psi_{jm'}(\mathbf{r}) \tag{5.138[3]}$$

From Eqs. (5.138[1],[2]) and (5.139), we have,

$$\mathcal{D}^j_{m'm}(\alpha\beta\gamma) = \exp(-im'\alpha)d^j_{m'm}(\beta) \exp(-im\gamma). \tag{5.141}$$

$\mathcal{D}^j$ is called the *rotation matrix*.

Wigner has derived the following expression for $d^j_{m'm}(\beta)$:

$$d^j_{m'm}(\beta) = \sum_{\kappa}(-1)^{\kappa} \frac{[(j+m)!(j-m)!(j+m')!(j-m')!]^{1/2}}{\kappa!(j-m'-\kappa)!(j+m-\kappa)!(\kappa+m'-m)!}$$

$$\times(\cos(\beta/2))^{2j-m+m'-2\kappa}(-\sin(\beta/2))^{m'-m+2\kappa} \tag{5.142}$$

where, κ takes positive integer values (including zero) consistent with the factorial notation. Also, for $m' \geq m$, we have the following more concise expression :

$$d^j_{m'm}(\beta) = \left\{\frac{(j-m)!(j+m')!}{(j+m)!(j-m')!}\right\}^{1/2} \cdot \frac{(\cos\beta/2)^{2j+m-m'}(-\sin\beta/2)^{m'-m}}{(m'-m)!}$$

$$\times 2F_1(m'-j, -m-j, m'-m+1, -\tan^2\beta/2), \tag{5.142a}$$

where ${}_2F_1(\,)$ is a hyper geometric function.

Properties of the Rotation Matrix

In the following, we will omit the arguments $(\alpha\beta\gamma)$ wherever their explicit appearance is not essential :

(i) Unitarity

The unitarity of $\mathcal{D}^j$ is expressed by

$$\sum_m \mathcal{D}^{j*}_{mm'} \mathcal{D}^j_{mm''} = \delta_{m'm''}. \tag{5.143a}$$

and

$$\sum_m \mathcal{D}^j_{m'm} \mathcal{D}^{j*}_{m''m} = \delta_{m'm''}. \tag{5.143b}$$

(ii) Symmetry

$$d^j_{m'm}(+\beta) = (-1)^{m'-m} d^j_{m'm}(-\beta), \tag{5.144a}$$

$$= d^j_{mm'}(-\beta) \tag{5.144b}$$

$$= (-1)^{m'-m} d^j_{-m'-m}(+\beta) \tag{5.144c}$$

$$= (-1)^{j+m} d^j_{-m'm}(\pi - \beta) \tag{5.144d}$$

$$= (-1)^{2j} d^j_{m'm}(\beta + 2\pi) \tag{5.144e}$$

These relationships follow from Eq. (5.142). From these and Eq. (5.141), we get

$$D^{j*}_{m'm}(\alpha\beta\gamma) = \mathcal{D}^j_{mm'}(-\gamma, -\beta, -\alpha) \tag{5.145a}$$

$$= (-)^{m'-m} \mathcal{D}^j_{-m'-m}(\alpha\beta\gamma), \tag{5.145b}$$

$$\mathcal{D}^j_{m'm}(\alpha, \beta + 4\pi, \gamma) = (-1)^{2j} \mathcal{D}^j_{mm'}(\alpha, \beta + 2\pi, \gamma) \tag{5.145c}$$

$$= \mathcal{D}^j_{m'm}(\alpha\beta\gamma). \tag{5.145d}$$

Equations (5.145c,d) show that, in the case of a system with half-integral spin, a rotation through an angle 2π about an axis changes the sign of the wavefunction of the system. Such a change, however, does not lead to any unphysical situation since only the absolute square of the wavefunction has a direct physical meaning and since all physical observables (that is, their matrix elements) are bilinear in the wavefunction [Eq. (2.117a)]. Therefore, property (5.145c) does not prevent the existence of half-integral spins.

Problem 5.20: Verify relationships (5.145 a-d).

Problem 5.21: Obtain the elements of the rotation matrix $\mathcal{D}^j(\alpha\beta\gamma)$.

(iii) Sum Rules

In the case of a system consisting of two particles with angular momenta j_1 and j_2, we have [Eqs. (5.68a) (5.84b)],

$$|j_1 m_1> |j_2 m_2> = \sum_j C^{j_1 j_2 j}_{m_1 m_2 m} |jm> . \tag{5.146}$$

If the system is now rotated through angles $(\alpha\beta\gamma)$, we have from Eq. (5.138[3]),

$$\sum_{\mu_1,\mu_2} \mathcal{D}^{j_1}_{\mu_1 m_1} \mathcal{D}^{j_2}_{\mu_2 m_2} |j_1\mu_1> |j_2\mu_2> = \sum_{j\mu} C^{j_1 j_2 j}_{m_1 m_2 m} \mathcal{D}^{j}_{\mu m} |j\mu> . \tag{5.147}$$

Writing,

$$|j_1\mu_1> |j_2\mu_2> = \sum_{j'} C^{j_1 j_2 j'}_{\mu_1 \mu_2 \mu'} |j'\mu'>$$

in Eq. (5.147) and then equating coefficient of $|j\mu>$ on either side, we get,

$$C^{j_1 j_2 j}_{m_1 m_2 m} \mathcal{D}^{j}_{\mu m} = \sum_{\mu_1 \mu_2} C^{j_1 j_2 j}_{\mu_1 \mu_2 \mu} \mathcal{D}^{j}_{\mu_1 m_1} \mathcal{D}^{j}_{\mu_2 m_2} . \tag{5.148}$$

Using properties (5.84a, b) of the C-coefficient, we get,

$$\mathcal{D}^{j}_{\mu m} = \sum_{\mu_1 m_1} C^{j_1 j_2 j}_{\mu_1 \mu_2 \mu} C^{j_1 j_2 j}_{m_1 m_2 m} \mathcal{D}^{j_1}_{\mu_1 m_1} \mathcal{D}^{j_2}_{\mu_2 m_2} . \tag{5.149a}$$

$$\mathcal{D}^{j_1}_{\mu_1 m_1} \mathcal{D}^{j_2}_{\mu_2 m_2} = \sum_j C^{j_1 j_2 j}_{m_1 m_2 m} C^{j_1 j_2 j}_{\mu_1 \mu_2 \mu} \mathcal{D}^{j}_{\mu m} ; \tag{5.149b}$$

where $m_2 = m - m_1 ; \mu_2 = \mu - \mu_1 .$

Eqs. (5.149a, b) are known as the *Clebsch-Gordon series*.

(iv) Relationship to Spherical Harmonics

Substituting for $\psi_{jm}(\mathbf{r})$ in Eq. (5.138[3]) by the spherical harmonics $Y_{lm}(\theta, \phi)$ which are eigenfunctions of the orbital angular momentum [Eq. 5.46)], we have,

$$\hat{R}(\alpha\beta\gamma) Y_{lm}(\theta, \phi) = Y_{lm}[\hat{R}^{-1}(\theta, \phi)] \equiv Y_{lm}(\theta', \phi')$$

$$= \sum_{m'=-l}^{+l} \mathcal{D}^{l}_{m'm}(\alpha\beta\gamma) Y_{lm'}(\theta, \phi). \tag{5.150}$$

Here, (θ, ϕ) are the spherical co-ordinates of a point in the physical system after rotation and (θ', ϕ') the spherical co-ordinates of the same point before rotation [see Fig. (5.8a, b)]. Also, $m'\hbar$ could be regarded as the component of the angular momentum along the space-fixed z-axis and $m\hbar$ the component along the z-axis of a frame fixed in the physical system ('body-fixed' frame).

Let P_1 and P_2 be two points on a unit sphere, with spherical co-ordinates (θ_1, ϕ_1) and (θ_2, ϕ_2), respectively, after the rotation and (θ'_1, ϕ'_1) and (θ'_2, ϕ'_2) before the rotation. Then, from Eq. (5.150), we have,

$$Y^*_{lm}(\theta'_1, \phi'_1) = \sum_{m'} \mathcal{D}^{l*}_{m'm}(\alpha\beta\gamma) Y^*_{lm'}(\theta_1, \phi_1), \tag{5.150a}$$

and

$$Y_{lm}(\theta'_2, \phi'_2) = \sum_{m''} \mathcal{D}^{l}_{m''m}(\alpha\beta\gamma) Y_{lm''}(\theta_2, \phi_2). \tag{5.150b}$$

Thus,

$$\sum_{m=-l}^{+l} Y^*_{lm}(\theta'_1, \phi'_1) Y_{lm}(\theta'_2, \phi'_2)$$

$$= \sum_{m', m''} \sum_{m} \mathcal{D}^{l*}_{m'm} \mathcal{D}^{l}_{m''m} Y^*_{lm'}(\theta_1, \phi_1) Y_{lm''}(\theta_2, \phi_2)$$

$$= \sum_{m'=-l}^{+l} Y^*_{lm'}(\theta_1, \phi_1) Y_{lm'}(\theta_2, \phi_2), \tag{5.151}$$

by Eq. (5.143b). That is,

$$K \equiv \sum_{m=-l}^{+l} Y^*_{lm}(\theta_1, \phi_1) Y_{lm}(\theta_2, \phi_2),$$

is invariant under rotations. This fact can be utilized to evaluate K. For, choosing the co-ordinate axes such that P_1 is on the z-axis ($\theta_1 = 0$) and P_2 in the zx-plane ($\phi_2 = 0$), we have,

$$K = \sum_{m} Y^*_{lm}(0, \phi_1) Y_{lm}(\theta_2, 0)$$

$$= \sqrt{\frac{2l+1}{4\pi}} Y_{l_0}(\theta, 0) = \left(\frac{2l+1}{4\pi}\right) P_l(\cos\theta),$$

by Eq. (5.63), and since,

$$Y_{lm}(0, \phi_1) = \delta_{m,o} \sqrt{\frac{2l+1}{4\pi}}, \qquad \text{[see Eq. (5.55b)]}.$$

Also, θ is the angle between the radial vectors to the points P_1 and P_2 (see Fig. 5.11a). Thus,

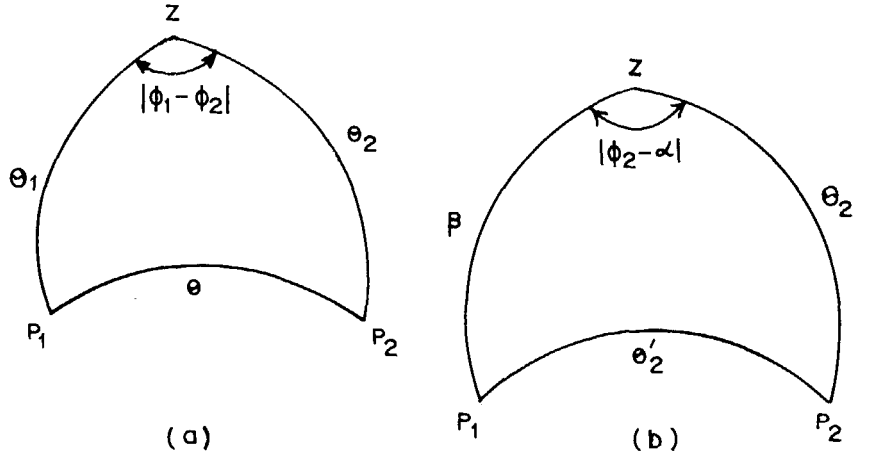


Fig. 5.11 The spherical triangles corresponding to : (a) Eq. (5.152), (b) Eq. (5.153b).

$$P_l(\cos\theta) = \left(\frac{4\pi}{2l+1}\right) \sum_{m=-l}^{+l} Y^*_{lm}(\theta_1, \phi_1) Y_{lm}(\theta_2, \phi_2). \tag{1.152}$$

Equation (5.152) holds for any spherical triangle (Fig. 5.11a) with sides $\theta_1, \theta_2, \theta$ and dihedral angle $|\phi_1 - \phi_2|$ opposite to θ.

If we put $\gamma = 0$ in Eq. (5.150b), then β and α would be, respectively, the polar and the azimuthal angles of oz' in the xyz-frame (Fig. 5.12). Let oz and oz' cut the unit sphere at the point z and P_1 respectively. Also choose the ox' and oy' axes such that P_2 lies in the $z'x'$-plane (that is, $\phi'_2 = 0$). Then, Eq. (5.150b) reduces to,

$$Y_{l_0}(\theta'_2, 0) = \sum_{m''} \mathcal{D}^l_{m''0}(\alpha\beta 0) Y_{lm''}(\theta_2, \phi_2). \tag{5.153a}$$

But P_1P_2z form a spherical triangle with sides β, θ_2 and θ'_2 and dihedral angle $|\phi_2 - \alpha|$ (Fig. 5.11b). Therefore, Eq. (5.152) can be applied, with $m = m''$, $\theta_1 = \beta$, $\theta = \theta'_2$ and $\phi_1 = \alpha$:

$$Y_{l_0}(\theta'_2, 0) = \sqrt{\frac{4\pi}{2l+1}} \sum_{m''=-l}^{+l} Y^*_{lm''}(\beta_\alpha) Y_{lm''}(\theta_2 \phi_2). \tag{5.153b}$$

Comparing Eqs. (5.153a, b), we have,

$$\mathcal{D}^l_{m0}(\alpha\beta 0) = \sqrt{\frac{4\pi}{2l+1}} Y^*_{lm}(\beta\alpha). \tag{5.154a}$$

$$= \sqrt{\frac{4\pi}{2l+1}} (-1)^m Y_{l-m}(\beta\alpha), \tag{5.154b}$$

where Eq. (5.59a) has been used.

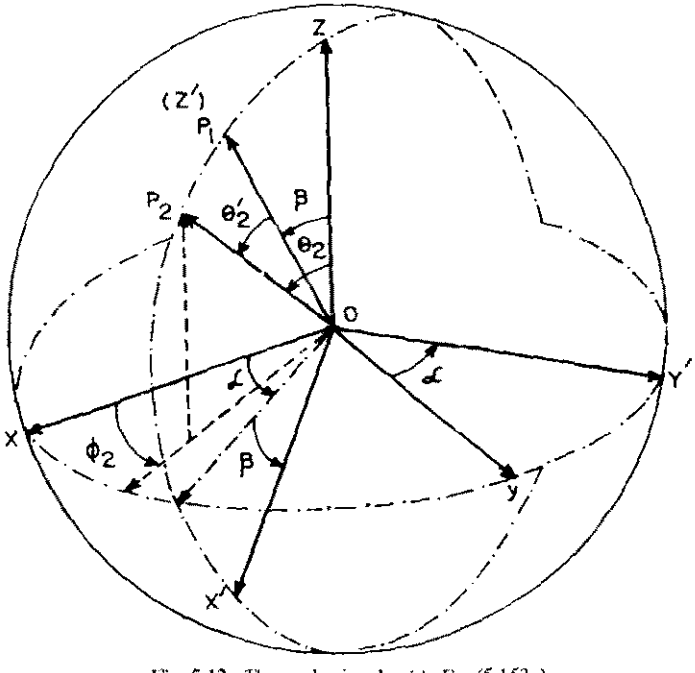


Fig. 5.12. The angles involved in Eq. (5.153a).

Also, from Eqs. (5.145a) and (5.154a), we get,

$$\mathcal{D}^{j}_{om}(0\beta\gamma) = \mathcal{D}^{j*}_{mo}(-\gamma, -\beta, 0)$$

$$= \sqrt{\frac{4\pi}{2l+1}}\, Y_{lm}(-\beta, -\gamma)$$

$$= \sqrt{\frac{4\pi}{2l+1}}(-1)^m Y^*_{lm}(\beta\gamma),$$

$$= \sqrt{\frac{4\pi}{2l+1}}\, Y_{l-m}(\beta\gamma), \qquad (5.154c)$$

using Eqs. (5.29a, b).

From Eqs. (5.154a) and (5.149b), we get,

$$Y_{l_1 m_1}(\theta,\phi) Y_{l_2 m_2}(\theta,\phi) = \sum_l \frac{1}{\sqrt{4\pi}} D_{l_1 l_2 l}\, C^{l_1 l_2 l}_{m_1 m_2 m}\, Y_{lm}(\theta,\phi). \qquad (5.155)$$

where $D_{l_1 l_2 l}$ is given by Eq. (5.87).

(v) Normalization

Writing, $\Theta \equiv (\alpha\beta\gamma)$, and $\int d\Theta \equiv \int_0^{2\pi} d\alpha \int_0^{\pi} \sin\beta d\beta \int_0^{2\pi} d\gamma,$ (5.156)

$$\int \mathcal{D}^{j}_{m'm}(\Theta) d\Theta = \int_0^{2\pi} e^{\,im'\alpha} d\alpha \int_0^{2\pi} e^{\,im\gamma} d\gamma \int_0^{\pi} d^{j}_{m'm}(\beta) \sin\beta d\beta.$$

$$= \delta_{m'0}\delta_{m,0}(2\pi)^2 \int_0^{\pi} d^{j}_{00}(\beta) \sin\beta d\beta,$$

since $\int_0^{2\pi} e^{-in\alpha} d\alpha = \delta_{n0}(2\pi).$

Now $d^{l}_{00}(\beta) = \mathcal{D}^{l}_{00}(0\beta 0) = \sqrt{\frac{4\pi}{2l+1}}\, Y_{l0}(\beta 0)$

$= P_l(\cos\beta)$, by Eqs. (5.154b) and (5.63),

so that, using Eq. (E. 31), we get,

$$\int \mathcal{D}^{j}_{m'm}(\Theta) d\Theta = 8\pi^2 \delta_{j0}\delta_{m0}\delta_{m'0} \qquad (5.157)$$

Using Eqs. (5.157) and (5.149b) and the properties of the C-coefficients, we derive,

$$\int D^{j_1}_{m_1 m_1}(\Theta)\mathcal{D}^{j_2}_{m_2 m_2}(\Theta)d\Theta = \left(\frac{8\pi^2}{2j_1+1}\right)\delta_{j_1 j_2}\delta_{m_1 m_2}\delta_{m'_1 m'_2}, \tag{5.158a}$$

and

$$\int D^{j_3}_{m_3 m_3}(\Theta)\mathcal{D}^{j_2}_{m_2 m_2}(\Theta)\mathcal{D}^{j_1}_{m_1 m_1}(\Theta)d\Theta\left(\frac{8\pi^2}{2j_3+1}\right)C^{j_1 j_2 j_3}_{m_1 m_2 m_3}C^{j_1 j_2 j_3}_{m_1 m_2 m_3} \tag{5.158b}$$

Problem 5.22: Using the relationship (5.154a) between the spherical harmonics and the rotation matrix, and Eqs. (5.157-158), establish the following relationships:

(a) $$\int Y_{lm}(\theta,\phi)d\Omega = \sqrt{4\pi}\,\delta_{lo}\delta_{mo}$$

(b) $$\int Y^*_{l_1 m_1}(\theta,\phi)Y_{l_2 m_2}(\theta,\phi)d\Omega = \delta_{l_1 l_2}\delta_{m_1 m_2}$$

(c) $$\int^* Y_{l_3 m_3} Y_{l_2 m_2} Y_{l_1 m_1} d\Omega = \frac{1}{\sqrt{4\pi}} D_{l_1 l_2 l_3} C^{l_1 l_2 l_3}_{m_1 m_2 m_3},$$

where $$d\Omega \equiv \sin\theta\, d\theta\, d\phi$$

(vi) Role as Angular Momentum Eigenfunction

The relationship of the $\mathcal{D}$-functions to the spherical harmonics suggests the possibility of interpreting the former as the eigenvectors of the angular momentum operator. We recall that $Y_{lm}(\beta\alpha)$ is the common eigenvector of $\hat{\mathbf{L}}^2$ and the component of $\hat{\mathbf{L}}$ that is conjugate to the angular variable α [Eqs. (5.44) to (5.46) and (5.42c)], namely $\hat{L}_z$. Therefore, from Eq. (5.154a), we have,

$$\hat{\mathbf{L}}_2\mathcal{D}^{l^*}_{mo}(\alpha\beta 0) = l(l+1)\hbar^2\mathcal{D}^{l^*}_{mo}(\alpha\beta 0), \tag{5.159^1}$$

$$\hat{L}_z\mathcal{D}^{l^*}_{mo}(\alpha\beta 0) = m\hbar\mathcal{D}^{l^*}_{mo}(\alpha\beta 0). \tag{5.160^1}$$

And from Eqs. (5.154c) and (5.145b), we get,

$$\hat{\mathbf{L}}^2 D^{l^*}_{om'}(0\beta\gamma) = l(l+1)\hbar^2\mathcal{D}^{l^*}_{om'}(0\beta\gamma), \tag{5.159^2}$$

$$\hat{L}_{z'}D^{l^*}_{om'}(0\beta\gamma) = m'\hbar\mathcal{D}^{l^*}_{om'}(0\beta\gamma), \tag{5.161^1}$$

where $\hat{L}_{z'}$ is the component of $\hat{\mathbf{L}}$ that is conjugate to the angular variable γ and hence is along the z'-axis.

Equations (5.159^1-161^1) suggest the generalisation:

$$\hat{\mathbf{J}}^2\mathcal{D}^{j^*}_{mk}(\alpha\beta\gamma) = j(j+1)\hbar^2\mathcal{D}^{j^*}_{mk}(\alpha\beta\gamma), \tag{5.159^3}$$

$$\hat{J}_z\mathcal{D}^{j^*}_{mk}(\alpha\beta\gamma) = m\hbar\mathcal{D}^{j^*}_{mk}(\alpha\beta\gamma), \tag{5.160^2}$$

$$\hat{J}_{z'}\mathcal{D}^{j^*}_{mk}(\alpha\beta\gamma) = k\hbar\mathcal{D}^{j^*}_{mk}(\alpha\beta\gamma). \tag{5.161^2}$$

Thus, $\mathcal{D}^{j*}_{mk}(\alpha\beta\gamma)$ is the common eigenvector of $\hat{\mathbf{J}}^2$ and the components of $\hat{\mathbf{J}}$ along the z and z'-axes. A physical interpretation of this eigenvector is the following: The $x'y'z'$-frame could be one fixed to a physical system. Then the Euler angles give the orientation of the body (the physical system) relative to the space-fixed frame xyz. If the Euler angles are time-varying, the physical system is a rotating body[19]. $m\hbar$ and $k\hbar$ are, respectively, the projections of the angular momentum on the space-fixed z-axis and the body-fixed z-axis, and $\mathcal{D}^{j*}_{mk}(\alpha\beta\gamma)$ is the wavefunction of the rotating system[20]. If $\Phi_{jmk}(\Theta)$ represents the normalized wavefunction, then, we have from Eq. (5.158a),

$$\Phi_{jmk}(\Phi) = \left(\frac{2j+1}{8\pi^2}\right)^{1/2} \mathcal{D}^{j*}_{mk}(\Theta). \tag{5.162}$$

The commutation of $\hat{J}_z$ and $\hat{J}_{z'}$ implied by Eqs. (5.160[2]) and Eqs. (5.161[2]) follows from the fact that the components of $\hat{\mathbf{J}}$ along the body-fixed (also referred to as 'intrinsic') frame cannot depend on the orientation of the space-fixed frame.

We note here the dual role of the $\mathcal{D}$-functions - as an operator and as a vector. This dual role is, in fact, a feature of certain types of operators called 'spherical tensors' which would be discussed in the next section.

5.7 SPHERICAL TENSORS

The $(2k + 1)$ operators, $\hat{T}^{(k)}_q$ for $q = -k, -k+1, \ldots +k$, are said to form the components of a *spherical tensor* of rank k if they transform under rotations like the spherical harmonic Y_{kq} of order k. That is, [see Eq. (5.150)].

$$\hat{T}^{(k)}_q(\mathbf{r}) \xrightarrow[(\alpha\beta\gamma)]{} \hat{T}^{(k)}_q(\mathbf{r}_1) = \Sigma_{q'} \mathcal{D}^{k}_{q'q}(\alpha\beta\gamma)\hat{T}^{(k)}_{q'}(\mathbf{r}). \tag{5.163[1]}$$

Here $\mathbf{r}_1$ is the position vector of a physical point in the rotated co-ordinate system (the 'body-fixed frame', of Section 5.5 (vi)) while $\mathbf{r}$ is the position vector of the same point in the original (space-fixed) frame. Now, a tensor is an entity which is defined primarily by its transformation properties under rotations. The adjective 'spherical' emphasizes the difference between a tensor of the type under consideration and a general (cartesian) tensor. This difference relates to a property called reducibility. This property is best understood through the example of a cartesian tensor of rank 2. Let T_{ij} represent the components of such a tensor in the 3-dimensional physical space. The total number of components is $3^2 = 9$. Each of these components transforms under rotations according to the formula.

---

19. This would imply that the body has a non-spherical shape, as the different orientations of a spherical body are indistinguishable.
20. For a more detailed treatment of this aspect, see Bohr, A. and Mottelson, B.R. *Nuclear Structure*, Vol. 1 (W.A. Benjamin, New York 1969). Section 1A-6. Note that the $\mathcal{D}$-function used by these authors is the complex conjugate of the one used in this book.

$$T_{ij}\overrightarrow{(\alpha\beta\gamma)}T'_{ij} = \sum_{kl} a_{ik}a_{jl}T_{kl}, \tag{5.164}$$

where the a_{ik} are the elements of an orthogonal matrix representing the rotation. Out of the nine components, we can form three groups of linear combinations of the components:

$$S = \frac{1}{3}\sum_i T_{ii}, \tag{5.165a}$$

$$V_k = \frac{1}{2}(T_{ij} - T_{ji}), \qquad i,j,k \text{ cyclic};$$

$$= \frac{1}{2}\in_{ijk} T_{ij}, \tag{5.165b}$$

$$A_{ij} = \frac{1}{2}(T_{ij} + T_{ji} - 2S\delta_{ij}), \tag{5.165c}$$

such that,

$$T_{ij} = A_{ij} + \in_{ijk} V_k + S\delta_{ij}. \tag{5.166}$$

Here $\in_{ijk}$ is the antisymmetric tensor introduced in Eq. (5.2b).

The peculiarity of the above three groups is that each of them transforms, under rotations, independently of the other two. In fact, S, being proportional to the trace of T, is invariant under rotations [see Eq. (A. 51)], and, hence, is a *scalar*, or tensor of rank 0. V_1, V_2 and V_3 are the three independent components of an antisymmetric, second rank tensor, and so transforms like a *vector* (tensor of rank 1). A is a traceless, symmetric tensor of rank 2, and has, therefore, five independent components which transform among themselves under rotations. Thus, from the viewpoint of their behaviour under rotations, each of the three groups has a status that is independent of the other two. That is, each of them is a tensor. They differ from the cartesian tensor T in that their components cannot be organised into smaller subgroups such that each of the subgroups transforms under rotations independently of the other subgroups. They are, therefore, called *irreducible tensors*. On the other hand, a tensor like T, whose components or linear combinations of the components, can be divided into two or more groups which transform under rotations among themselves, is a *reducible tensor*.

The spherical tensors are nothing but the irreducible tensors that result from the grouping of the components of a general (cartesian) tensor as explained above. Of course, the components given by Eqs. (5.165b, c) are not the spherical components that transform like the components of the spherical harmonics. We can deduce the spherical components of V and of the symmetric tensor A with the help of the expressions for Y_{1m} and Y_{2m} given in Eq. (5.63). We have [see also Eq. (4.107)],

$$rY_{10} = \sqrt{\frac{3}{4\pi}}z; \; rY_{1\pm1} = \sqrt{\frac{3}{4\pi}}\left\{\mp\frac{1}{\sqrt{2}}(x \pm iy)\right\},$$

which show that the spherical components of the vector **V** are given by,

$$T_0^{(1)} \equiv V_0 = V_3 = V_z,$$

$$T_{\pm 1}^{(1)} \equiv V_{\pm 1} = \mp \frac{1}{\sqrt{2}}(V_1 \pm iV_2) = \mp \frac{1}{\sqrt{2}}(V_x + iV_y). \tag{5.167}$$

Similarly, from

$$r^2 Y_{20} = \sqrt{\frac{5}{16\pi}}(2z^2 - x^2 - y^2),$$

$$r^2 Y_{2\pm 1} = \sqrt{\frac{5}{16\pi}}\{\mp\sqrt{6}(xz \pm iyz)\},$$

$$r^2 Y_{2\pm 2} = \sqrt{\frac{5}{16\pi}}\left\{\sqrt{\frac{3}{2}}(x^2 - y^2 \pm 2ixy)\right\},$$

we get,

$$T_0^{(2)} = 2A_{33} - A_{11} - A_{22},$$

$$T_{\pm 1}^{(2)} = \mp\sqrt{6}\,(A_{13} \pm iA_{23}), \tag{5.168}$$

$$T_{\pm 2}^{(2)} = \sqrt{\frac{3}{2}}(A_{11} - A_{22} \pm 2iA_{12}).$$

It turns out that the irreducible tensors that are derived by the reduction of a general tensor, have invariably odd numbers of components. This means that the spherical tensors are of integral rank (as also implied by their definition in terms of the spherical harmonics). However, the transformation law (5.163) permits spherical tensors of half-integral rank also, as the $\mathcal{D}$-matrix can be of half-integral order. Such tensors[21] are called *spinors*, and represent wavefunctions of particles with half-integral spin. Irreducible tensor operators representing interactions between physical systems are invariably of integral order.

The importance of spherical tensors in the theory of angular momentum stems from the fact that the spherical tensor $\hat{T}_q^{(k)}$, like the spherical harmonic Y_{kq}, is associated with an angular momentum k with z-component q (That is, $|\mathbf{k}| = \sqrt{k(k+1)}\hbar$; $k_z = q\hbar$). Thus, if $|00>$ represents a state with zero angular momentum ($j = 0$; $m = 0$), then, $\hat{T}_q^{(k)}|00>$ should be a state with angular momentum k and z component q:

$$\hat{T}_q^{(k)}|00> \propto |kq>, \tag{1.169}$$

Similarly, $\hat{T}_q^{(k)}|jm>$ would yield a state $|j'm'>$ with $m' = m + q$ and $|j - k| \leq j' \leq (j + k)$. This fact is better expressed by the equation,

$$<j'm'|\hat{T}_q^{(k)}|jm> = f(j, j', k)C_{mqm'}^{jkj'}, \tag{5.170[1]}$$

---

21. For a more detailed discussion on these tensors, see de-Shalit, A. and Talmi, I. *Nuclear Shell Theory* (Academic Press, New York 1963), Section 11.

where, $f()$ is independent of the projection quantum numbers [see, Eq. (5.170^2) below].

Commutation Relations with the Components of $\hat{\mathbf{J}}$

Since the $\hat{T}^{(k)}_q$ are operators, Eq. (5.163^1) can also be written as [see Eq. (5.129)],

$$\hat{R}(\Theta)\hat{T}^{(k)}_q\hat{R}^{-1}(\Theta) = \sum_{q'} \mathcal{D}^k_{q'q}(\Theta)\hat{T}^{(k)}_{q'}, \qquad (5.163^2)$$

where $$\Theta \equiv (\alpha, \beta, \gamma)$$

Substituting for $\hat{R}(\Theta)$ in Eq. (5.163^2) the infinitesimal rotation operator given by Eq. (5.128^3) and using the relation,

$$\mathcal{D}^k_{q'q}(\mathbf{n}, \delta\theta) = < kq' \mid \hat{R}_{\mathbf{n}}(\delta\theta) \mid kq >$$

$$= \delta_{qq'} - \frac{1}{\hbar}\delta\theta < kq' \mid (\mathbf{n}\cdot\hat{\mathbf{J}}) \mid kq >, \qquad (5.171)$$

we get,

$$[(\mathbf{n}\cdot\hat{J}), \hat{T}^{(k)}_q] = \sum_{q'} < kq' \mid (\mathbf{n}\cdot\hat{\mathbf{J}}) \mid kq > \hat{T}^{(k)}_{q'}, \qquad (5.172)$$

or,

$$[\hat{J}_{\pm}, \hat{T}^{(k)}_q] = \hbar\{(k \mp q)(k \pm q + 1)\}^{1/2}\hat{T}^{(k)}_{q\pm 1}, \qquad (5.172a)$$

$$[\hat{J}_z, \hat{T}^{(k)}_q] = \hbar q\, \hat{T}^{(k)}_q. \qquad (5.172b)$$

Eqs. (5.172a, b) are equivalent to the Eq. (5.163^2), and could, therefore, be regarded as an alternative definition of the spherical tensors.[22]

Problem 5.23: Obtain the commutation relations (5.16a, b) from Eqs. (5.172a, b).

Product of Spherical Tensors

Scalar product:

$$(\hat{T}^{(k)}.\hat{U}^{(l)}) = \left(\sum_q (-1)^q \hat{T}^{(k)}_q \hat{U}^{(k)}_{-q}\right)\delta_{kl} \qquad (5.173)$$

Tensor product:

$$[\hat{T}^{(r)} \times \hat{U}^{(s)}]^{(t)}_m = \sum_q \hat{T}^{(r)}_q \hat{U}^{(s)}_{m-q} C^{rst}_{qm-qm} \qquad (5.174)$$

From the C-coefficient, it is clear that the rank of the tensor product can vary from $|r-s|$ to $(r+s)$. In the case of four commuting spherical tensors $\hat{F}^{(r)}, \hat{G}^{(r)}, \hat{H}^{(s)}$ and $\hat{M}^s$, we have,

$$(\hat{F}^{(r)}.\hat{G}^{(r)})(\hat{H}^{(s)}.\hat{M}^{(s)}) = (-1)^{r+s}\sum_t (-1)^t ([\hat{F}^{(r)} \times \hat{H}^{(s)}]^t.[\hat{G}^{(r)} \times \hat{M}^{(s)}]^t). \qquad (5.175)$$

---

22. This definition is due to Racah (Reference cited in foot note 11).

If $\hat{T}^{(r)}$ and $\hat{U}^{(s)}$ are spherical tensors operating on different spaces, then we can define a double tensor by

$$\hat{T}^{(r)}_{\mu} U^{(s)}_{\nu} = \sum_{t} C^{rst}_{\mu\nu(\mu+\nu)} [\hat{T}^{(r)} \times \hat{U}^{(s)}]^{t}_{\mu+\nu}. \tag{5.176}$$

This is the inverse of Eq. (5.174).

Matrix Elements: The Wigner-Eckart Theorem

Eq. (5.170[1]) is referred to as the Wigner-Eckart theorem. A more convenient (and conventional) notation is,

$$< j_2 m_2 \mid \hat{T}^{(k)}_{q} \mid j_1 m_1 > = C^{j_1 k j_2}_{m_1 q m_2} \frac{< j_2 \| \hat{T}^{(k)} \| j_1 >}{\sqrt{2j_2+1}}, \tag{5.170[2]}$$

where, the double-barred matrix element is known as the *reduced matrix element*. The theorem basically divides the matrix element into two factors—one (the C-coefficient) depending on the geometry (that is, the relative orientation of the physical and the co-ordinate systems) and the other (the reduced matrix element) on the intrinsic properties of the tensor $\hat{T}^{(k)}$. A proof of the theorem is given below:

$$< j_2 m_2 \mid \hat{T}^{(k)}_{q} \mid j_1 m_1 > \equiv \int \psi^{*}_{j_2 m_2}(\mathbf{r}) \hat{T}^{(k)}_{q} \psi_{j_1 m_1}(\mathbf{r}) d^3\mathbf{r}$$

$$= \int \{\hat{R}(\Theta)\psi_{j_2 m_2}(\mathbf{r})\}^{*} \{\hat{R}\hat{T}^{(k)}_{q}\hat{R}^{-1}\} \{\hat{R}\psi_{j_1 m_1}(\mathbf{r})\} d^3\mathbf{r}$$

$$= \sum_{m_1' m_2' q'} \int \psi^{*}_{j_2 m_2'}(\mathbf{r}) \hat{T}^{(k)}_{q'} \psi_{j_1 m_1'}(\mathbf{r}) d^3\mathbf{r} \mathcal{D}^{j_2 *}_{m_2' m_2}(\Theta) \times \mathcal{D}^{k}_{q' q}(\Theta) \mathcal{D}^{j_1}_{m_1' m_1}(\Theta),$$

by Eqs. (5.138[3]) and (5.163[2]),

$$= \sum_{m_1' m_2' q'} < j_2 m_2' \mid \hat{T}^{(k)}_{q'} \mid j_1 m_1' > \mathcal{D}^{j_2 *}_{m_2' m_2}(\Theta) \times \mathcal{D}^{k}_{q' q}(\Theta) D^{j_1}_{m_1 m_1}(\Theta). \tag{5.177}$$

Integrating both sides over the Euler angles (Θ) and making use of Eqs. (5.156) and (5.158b), we get,

$$< j_2 m_2 \mid \hat{T}^{(k)}_{q} \mid j_1 m_1 > \times 8\pi^2 = \sum_{m_1' m_2' q'} < j_2 m_2' \mid \hat{T}^{(k)}_{q'} \mid j_1 m_1' > \times \frac{8\pi^2}{2j_2+1} C^{j_1 k j_2}_{m_1 q m_2} C^{j_1 k j_2}_{m_1' q' m_2'}.$$

That is,

$$< j_2 m_2 \mid \hat{T}^{(k)}_{q} \mid j m_1 > = C^{j_1 k j_2}_{m_1 q m_2} \frac{1}{\sqrt{2j_2+1}} \{ \sum_{m_1' m_2' q'} \frac{1}{\sqrt{2j_2+1}} \times C^{j_1 k j_2}_{m_1' q' m_2'} < j_2 m_2' \mid \hat{T}^{(k)}_{q'} \mid j_1 m_1' > \}. \tag{5.170[3]}$$

The quantity in the curly bracket is, obviously, independent of the projection quantum numbers and could, therefore, be written as

$$\sum_{m_1' m_2' q'} \frac{1}{\sqrt{2j_2+1}} \cdot C^{j_1 k j_2}_{m_1' q' m_2'} < j_2 m'_2 \mid \hat{T}^{(k)}_{q'} \mid j_1 m'_1 >$$

$$= < j_2 \| \hat{T}^{(k)} \| j_1 > . \qquad (5.178)$$

Substitution of (5.178) in (5.170^3) leads to Eq. (5.170^2).

In the case of a scalar operator, $\hat{S} \equiv \hat{T}^{(o)}_o$, we have from Eq. (5.170^2) and the relation $C^{j0j'}_{mom'} = \delta_{jj'}\delta_{mm'}$,

$$< j'm' \mid \hat{S} \mid jm > = \frac{1}{\sqrt{2j+1}} < j \| \hat{S} \| j > \delta_{jj'}\delta_{mm'} \qquad (5.179)$$

The unit operator $\hat{1}$ is a scalar. Hence from Eq. (5.179), we have,

$$< j' \| \hat{1} \| j > = \sqrt{2j+1}\, \delta_{jj'}. \qquad (5.180)$$

Problem 5.24: Deduce the following relationships:

(a) $< j' \| \hat{J} \| j > = \hbar \sqrt{j(j+1)(2j+1)}\, \delta_{jj'}$

(b) $< l' \| C^{(L)} \| l > = (-1)^l D_{l'lL}$,

where, $C^{(L)}_M = \left(\frac{4\pi}{2L+1} \right)^{1/2} Y_{LM}$, and $D_{l'lL}$

is the angular momentum coupling coefficient defined by (5.87).

Given below are some of the most commonly used formulae involving the reduced matrix elements of spherical tensors. The derivation of these relationships are straightforward, though tedious[23].

Consider a two-particle system with a wavefunction $\mid j_1 j_2 jm >$ in the notation of Eq. (5.68a). Let $\hat{T}^{(k)}$ and $\hat{U}^{(k)}$ be irreducible tensor operators defined on the Hilbert space of particle 1 and particle 2, respectively. Then,

$$< j_1' j_2' j_1' \| (\hat{T}^{(k)} \| j_1 j_2 j > = \{(2j+1)(2j'+1)\}^{1/2} W(j_1' j_2 k j : j' j_1)$$

$$\times < j_1' \| \hat{T}^{(k)} \| j_1 > \delta_{j_2 j_2'}. \qquad (5.181a)$$

$$< j_1' j_2' j_1' \| \hat{U}^{(k)} \| j_1 j_2 j > = \{(2j+1)(2j'+1)\}^{1/2} W(j_1 j' j_2 k : j_2' j)$$

$$\times < j_2' \| \hat{U}^{(k)} \| j_2 > \delta_{j_1 j_1'}. \qquad (5.181b)$$

$$< j_1' j_2' j' m' \mid (\hat{T}^{(k)} . \hat{U}^{(k)}) \mid j_1 j_2 jm >$$

$$= \delta_{mm'}\delta_{jj'} (-1)^{j_1' + j_2 - j} W(j_1' j_2' j_1 j_2 : jk) < j_1' \| T^{(k)} \| j_1 > \times < j_2' \| \hat{U}^{(k)} \| j_2 > . \qquad (5.182a)$$

23. The derivation of some of these can be found in Section 24 of Ref. 2 listed at the end of this chapter.

$$< j_1' j_2' j' || [\hat{T}^{(r)} \times \hat{U}^{(s)}]^{(t)} || j_1 j_2 j >$$

$$= \{(2j+1)(2j'+1)(2t+1)\}^{1/2} < j_1' || \hat{T}^{(r)} || j_1 >$$

$$\times < j_2' || \hat{U}^{(s)} || j_2 > \begin{Bmatrix} j_1' & j_2' & j' \\ j_1 & j_2 & j \\ r & s & t \end{Bmatrix}. \qquad (5.182b)$$

If $\hat{\mathbf{V}}$ is an arbitrary vector operator (that is, a spherical tensor of rank 1), then,

$$< j'm' | \hat{\mathbf{V}} | jm > = \delta_{jj'} \frac{< jm' | \hat{\mathbf{J}}(\hat{\mathbf{J}} \cdot \hat{\mathbf{V}}) | jm >}{j(j+1)\hbar^2}$$

$$= \delta_{jj'} \frac{< jm' | \hat{\mathbf{J}} | jm > < j || (\hat{\mathbf{J}} \cdot \hat{\mathbf{V}}) || j >}{j(j+1)\hbar^2} \qquad (5.183)$$

This equation is known as the *projection theorem for first rank tensors*[24]. In the case of a component $\hat{V}_q$ of $\hat{\mathbf{V}}$ the equation reduces to [in view of Eq. (5.170[2]) and Problem (5.24a)],

$$< j'm' | \hat{V}_q | jm > = \delta_{jj'} \frac{< jm' | \hat{J}_q | jm > < j || (\hat{\mathbf{J}} \cdot \hat{V}) || j >}{j(j+1)\hbar^2} \qquad (5.183a)$$

Trace of a Spherical Tensor

This is defined as,

$$Tr(\hat{T}_q^{(k)}) = \sum_{jm} < jm | \hat{T}_q^{(k)} | jm >$$

$$= \sum_{jm} C^{jkj}_{mqm} \frac{< j || \hat{T}^{(k)} || j >}{\sqrt{2j+1}}$$

$$= \delta_{qo} \delta_{ko} \sum_j \sqrt{2j+1} < j || T^{(o)} || j >, \qquad (5.184)$$

where use has been made of the relation $C^{joj}_{mom} = 1$, and Eqs. (5.83c) and (5.84a).

Thus, the trace of a spherical tensor of non-zero rank, is zero.

Problem 5.25: Deduce the above fact from the commutation relations (5.172 a, b).

5.8 CONSEQUENCES OF QUANTIZATION

The most important consequence of the quantization of angular momentum, is that the operators representing the components do not commute. The uncertainty principle, then, makes it impossible to measure accurately more than one component of angular momentum in any particular quantum state. According to

24. See Ref. 2, Section 20.

Eq. (3.28) [see, also Eq. (3.28a)], the uncertainties ΔJ_x and ΔJ_y in the values of J_x and J_y in the state $| jm >$ are given by the relation,

$$\Delta J_x \cdot \Delta J_y \geq \frac{1}{2} \mid \hbar < jm \mid \hat{J}_z \mid jm > \mid$$

$$\geq \frac{1}{2} \mid m \mid \hbar^2 \tag{5.185}$$

It follows as a corollary of this result that the angular momentum vector cannot be exactly parallel to the axis of quantization (the z-axis). For, if it could, the uncertainties in J_x and J_y would be zero (since J_x and J_y themselves would be exactly zero). In fact, it is possible to show (see Problem 3.5) that the relation (5.185) demands that an angular momentum whose maximum projection on the z-axis is $j\hbar$, should be of length $\sqrt{j(j+1)}\hbar$. Also, $< \hat{J}_x > = < \hat{J}_y > = 0$, so that **J** *precesses* around the z-axis. Consequently, the orbit of the particle would not lie in a single plane.

Another consequence of quantization is that the measured values of the angular momentum along a reference axis has to be an integral multiple of $(\hbar/2)$. This is a consequence of the particular commutation relations (5.7a) satisfied by the components of the angular momentum operator. It means that, when there is a preferred direction, such as a magnetic field in the case of a charged particle, the orientation of the angular momentum vector is quantized with respect to this direction; only certain discrete orientations being allowed.

That a general angular momentum can have half-integer values ($j = 0, 1/2, 1, 3/2, ...$), while the orbital angular momentum has only integer values ($l = 0, 1, 2, ...$), may be taken to imply the existence of a half-integral *intrinsic* angular momentum (Spin). Historically, the hypothesis of spin (see footnote 3) preceded the theory of angular momentum.

REFERENCES

1. EDMONDS, A.R. *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton 1957).
2. ROSE, M.E. *Elementary Theory of Angular Momentum* (John Wiley, New York 1961).
3. MESSIAH, A. *Quantum Mechanics*, Vol. II (North-Holland, Amsterdam 1961), Chapter XIII.
4. FRENCH, J.B. *Technical Report No. IX, Part I* (SMS Radiation Laboratory, University of Pittsburg, 1958).
5. ROTENBERG, M., BIVINS, R., METROPOLIS, N. and WOOTEN, JR. J.K. *The 3-j and 6-j symbols* (The Technology Press, M.I.T., Massachusetts, 1959).
6. HOWELL, K.M. *Research Report No.* 59-1 (University of Southampton, 1959).
7. KENNEDY, J.M. and CLIFF, M.J., *Transformation Coefficients between LS- and jj-coupling* (Atomic Energy Commission of Canada, 1955).
8. ARIMA, A., HORIE, H. and TANAKA, Y. *Progress of Theoretical Physics*, **11**, 143 (1954).

CHAPTER 6

INVARIANCE PRINCIPLES AND CONSERVATION LAWS

In the course of the dynamical development of a physical system, though many of the observables change their values, there might be one or more physical observables which do not change with time. Such observables are called *constants of motion*, and the principle embodying their constancy in time are referred to as *conservation laws*. Thus, in the case of a free particle, the linear momentum is a constant of motion, whereas for a system moving under the influence of a spherically symmetric field, the linear momentum varies with time but both the angular momentum and the energy are constants of motion. It will be shown below that the constancy of a physical observable implies, or rather stems from, the invariance of the Hamiltonian of the physical system under a certain *symmetry operation*. Among the various symmetries, the *geometrical symmetries* associated with space and time are of special importance. These are the symmetries arising from the homogeneity of space and time (which results in the invariance of the Hamiltonian under translations in space and time), the isotropy of space[1] (leading to the rotational invariance) and the invariance of the Hamiltonian under inversions of space and time. Then there are the *dynamical symmetries* which are associated with particular features of the interaction involved. For example, the $(2l + 1)$-fold degeneracy of the energy associated with a particular angular momentum l, is a feature of motion in a central field, and arises from the cylindrical symmetry of a central field[2]. Similarly, the l-degeneracy discussed in connection with the energy levels of the hydrogen atom (See Eqs. (4.117), (4.116) and (4.128)) is a consequence of a special symmetry of the Coulomb interaction[3]. In this chapter, we will confine ourselves to a discussion of the space-time symmetries and the associated conservation laws. We will first establish the relationship between a conservation law and a symmetry operation, and then discuss each of the symmetry operations in more detail.

---

1. Time being one-dimensional, we cannot talk of rotations in time.
2. The axis of the cylinder is perpendicular to the angular momentum vector. In a scattering process, for example, it is the incident direction.
3. For a more detailed discussion of this point in particular and of the dynamical symmetries in general, see L.I. Schiff, *Quantum Mechanics* (McGraw-Hill, 1968), III Edition, Section 30.

6.1 SYMMETRY AND CONSERVATION LAWS

From Eq. (4.25), we have, $\frac{d\langle\hat{A}\rangle}{dt} = 0$, if,

$$[\hat{A}, \hat{H}] = \hat{0}. \tag{6.1}$$

That is, an observable is a constant of motion if the corresponding operator commutes with the Hamiltonian. Now, corresponding to the Hermitian operator $\hat{A}$, we can define a unitary operator $\hat{U}_A$ [See Eq. (2.63)]:

$$\hat{U}_A(\epsilon) = e^{-i(\epsilon \hat{A})}, \tag{6.2}$$

where, ϵ is a real parameter, which is a scalar or a vector according as $\hat{A}$ is a scalar or a vector operator. $(\epsilon \hat{A})$ would be the scalar formed by taking the product of ϵ and $\hat{A}$. Eq. (6.1) implies the relationship (provided $\epsilon \hat{H} = \hat{H} \epsilon$),

$$\hat{U}_A \hat{H} = \hat{H} \hat{U}_A. \tag{6.3a}$$

or

$$\hat{U}_A \hat{H} \hat{U}_A^\dagger = \hat{H}. \tag{6.3b}$$

But, Eq. (6.3b) represents a unitary transformation. Hence we have the result that, if the observable corresponding to the operator $\hat{A}$ is conserved during the motion of the system, then the Hamiltonian of the system is invariant under the unitary transformation generated by $\hat{A}$. Such transformations (those that leave the Hamiltonian invariant) are called *symmetry transformations*. Thus, a conservation law invariably implies the existence of a symmetry transformation for the system. The converse is, however, not necessarily true, as we will presently see.

A symmetry transformation, in addition to leaving the Hamiltonian invariant, is characterised by the following properties:

(*ST*1) It preserves the Hermitian character of an operator. This ensures that observables remain observables under the transformation.

(*ST*2) It conserves probabilities. This means that the absolute value of the scalar product of a pair of vectors remains invariant under the transformation (Remember that probability is proportional to the square of the absolute value of the scalar product).

Both these properties imply that a symmetry transformation could be either unitary or *antiunitary* [See Eqs. (2.60)[4] and (B.11)]. The later possibility arises when $\hat{A}$ in Eq. (6.1) is both antilinear and unitary (that is, satisfies Eq. (B.1) as well as the condition $\hat{A}^\dagger = \hat{A}^{-1}$). Since an antilinear operator cannot represent a physical observable (as it does not preserve the principle of superposition), symmetry transformations corresponding to antiunitary transformations, do not lead to any conservation laws. Nevertheless, such transformations are of importance because of the *selection rules* they provide, which enable us to tell why transitions between certain states are allowed while those between certain others are forbidden.

4. The derivation leading to Eq. (2.60) is valid even when $\hat{U}$ is antilinear.

Among the space-time symmetries, it turns out that the time reversal operation is associated with an antiunitary operator. The other four symmetries correspond to unitary transformations.

Symmetry and Degeneracy

From Eq. (6.3a), we see that, if $|u_k>$ is an eigenvector of $\hat{H}$ belonging to the eigenvalue E_k, then,

$$\hat{H}\{\hat{U}_A \mid u_k\rangle\} = E_k\{\hat{U}_A \mid u_k\rangle\}, \tag{6.4}$$

so that, $\hat{U}_A \mid u_k\rangle$ is also an eigenvector of $\hat{H}$ belonging to the same eigenvalue. If $U_A \mid u_k\rangle$ is linearly independent of $\mid u_k\rangle$, then the energy eigenvalue E_k is *degenerate* (see, *Commuting Operators*, Section 2.2). Thus, degeneracy of the energy eigenvalues is another consequence of the invariance of the Hamiltonian under a symmetry transformation. Conversely, degeneracy of the energy eigenvalues implies the existence of an operator (not necessarily an observable) that commutes with the Hamiltonian and thereby points to the existence of a symmetry transformation for the system.

6.2 THE SPACE-TIME SYMMETRIES

As we have already stated, the space-time symmetries refer to the symmetries associated with geometrical operations like displacement, rotation and inversion (reflection at the origin) in space and time. For example, the assumption that space is homogeneous, requires that the Hamiltonian of (as well as the probabilities associated with) a physical system be invariant under displacement, or translation, in space. We will see that such a translational invariance implies the conservation of the linear momentum of the system. Similarly, the invariance under rotations has its origin in the isotropy of space (the fact that space has the same properties in all directions). The importance of these geometrical symmetries associated with space and time for physics would be clear if we think of the situation if these symmetries were absent:

Imagine an experiment, say for measuring the cross-section for the scattering of neutrons by protons, being performed at two laboratories, one at Madras and the other at Delhi. The expectation that the measurements at the two laboratories will yield the same result, within experimental errors, if performed under identical laboratory conditions, is based on the presumption that the differences in the location and the orientation (arising from the curvature of the Earth) of the two laboratories have no effect on the outcome of the experiment (that is, on the physics of the problem). Similarly, the fact that the repetition of the experiment will lead to the same result, depends on the independence of the outcome on the time at which the experiment is performed; that is, on the homogeneity of time. Thus, the reproducibility (and through it, the verifiability) of an experimental result at different locations and time, which is basic to the philosophy of all experimental sciences and which permits the formulation of laws with a universal validity, stems from the homogeneity of space and time and the isotropy of space.

Of course, even in the absence of these properties of space and time, it might be possible to arrive at laws of a general nature; but this would require a knowledge of the precise manner in which a displacement in space or time or a rotation in space affects the observations. In fact, we do have to cope up with local inhomogeneity and anisotropy, as in the neighbourhood of the earth. But these we are able to attribute to external (that is, not related to space and time) causes, such as the gravitational field in the case of the earth, and maintain that space is really homogeneous and isotropic everywhere[5]. It is clear that the formulation of laws of physics based on observations would have been a far more difficult task than it is now, if space and time were devoid of the geometrical symmetries under discussion.

We should here make a distinction between the *continuous symmetries* that are associated with displacement in space and time and rotation in space, and the *discrete symmetries*[6] that are associated with inversion of space and time. Whereas the former have universal validity (that is, they hold good for all known interactions[7]), the case of the latter is not so clear. For example, it is known that space-inversion invariance is not valid in the case of weak interactions. Time-reversal invariance also seems to be violated in certain reactions such as the decay of neutral kaons.[8]

We will now discuss the nature and properties of the operators, as well as the conservation laws, that are associated with each of the space-time symmetries.

6.2A Displacement in Space: Conservation of Linear Momentum

A displacement in space can be described either as a displacement of the co-ordinate system with the physical system fixed (the *passive* point of view) or as a displacement of the physical system with the co-ordinate system fixed (the *active* point of view). Obviously, a displacement of the physical system through the vector ρ, is equivalent to a displacement of the co-ordinate system through $-\rho$.

---

5. Alternatively, we should count the earth (the cause of the local inhomogeneity and anisotropy) also as part of the physical system. Then for such a combined system (which would form a *closed* system), space and time would be homogeneous and space isotropic.
6. These were introduced in quantum mechanics by E.P. Wigner during the early 1930's. [E.P. Wigner, *Gottinger Nachr*, **31**, 546 (1932); see Also, E.P. Wigner, *Group Theory and its Application in the Quantum Mechanics of Atomic Spectra* (Academic Press, New York 1959) Chapter 26.]
7. The presently-known interactions (forces) in nature are divided into four classes. These, in the order of decreasing strength, are: Strong (1), electromagnetic ($\sim 10^{-2}$), weak ($\sim 10^{-23}$) and gravitational ($\sim 10^{-39}$). Of these, the strong and the weak interactions are of extremely short range ($\sim 10^{-13}$) and, therefore, are dominant in the case of nuclear and subnuclear particles. Both the electromagnetic and the gravitational interactions are of long range, but the occurrence of two types of electric charges (positive and negative) and the electrical neutrality of macroscopic bodies, diminish the domain of dominance of electromagnetic interaction to that of atoms and molecules
8. The suggestion that space-inversion invariance may not hold good in the case of the weak interaction, was first put forward by T.D. Lee and C.N. Yang [*Physical Review*, **104**, 254 (1956)], and experimentally verified by C.S. Wu et. Al [*Physical Review*, **105**, 1413 (1957)].

Therefore, one is free to choose either of the viewpoints. We will adopt the active point of view in the case of the continuous symmetries, and the passive point of view in the case of the discrete symmetries.

Now, a continuous transformation should evolve from the identity transformation and should, therefore, be unitary. Let $\hat{\alpha}$ represent the Hermitian operator that is the generator of the infinitesimal displacement $\delta\rho$ (see Fig. 6.1) of the physical system. The unitary operator corresponding to this displacement, according to Eq. (6.2), is given by

$$\hat{U}_{\alpha}(\delta\rho) \approx \hat{1} - i(\delta\rho \cdot \hat{\alpha}). \tag{6.5^1}$$

The change in the wavefunction at the space point $\mathbf{r}$, resulting from this transformation is given (in the co-ordinate representation) by [(of. Eq. (5.127^2)],

$$\hat{U}_{\alpha}(\delta\rho)\psi(\mathbf{r}) = \psi'(\mathbf{r}) = \psi(\mathbf{r} - \delta\rho), \tag{6.6^1}$$

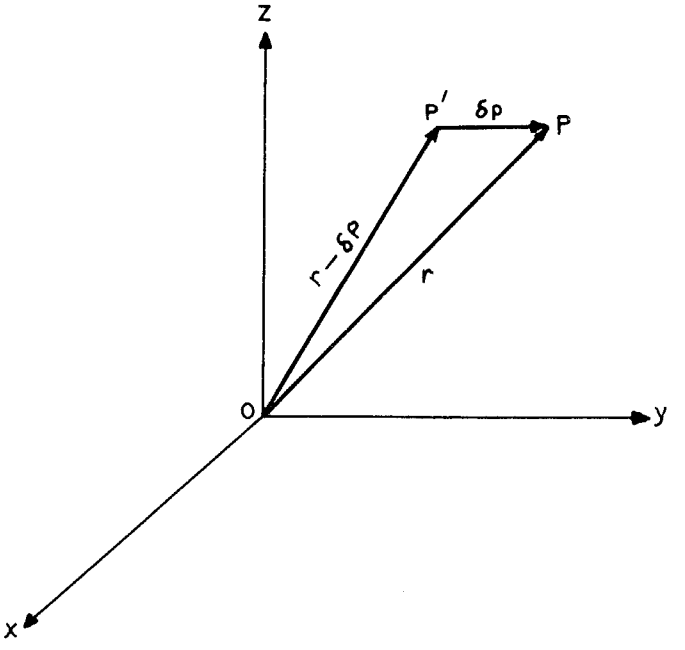


Fig. 6.1 Displacement $\delta\rho$ in space of a physical system

which means that the wavefunction at the point P after the displacement is the same as the wavefunction at P' (where P was) before the displacement. Alternatively, one could say that if the original (undisplaced wavefunction has a maximum at $\mathbf{r}$, the displaced wavefunction has a maximum at $\mathbf{r} + \delta\rho$. Using Eq. (6.5^1) and the Taylor series expansion of $\psi(\mathbf{r} - \delta\rho)$ around $\mathbf{r}$, Eq. (6.6^1) can be written as

$$[\hat{1} - i(\delta\rho \cdot \alpha)]\psi(\mathbf{r}) \approx [\hat{1} - (\delta\rho \cdot \nabla)]\psi(\mathbf{r})$$

$$\approx \left[\hat{1} - \frac{i}{\hbar}(\delta\rho \cdot \hat{\mathbf{p}})\right]\psi(\mathbf{r}), \tag{6.6^2}$$

where, $\hat{\mathbf{p}} = -i\hbar\nabla$, is the momentum operator.

Thus, $$\hat{\alpha} = \hat{\mathbf{p}}/\hbar; \tag{6.7}$$

so that, the unitary operator corresponding to the infinitesimal displacement $\delta\rho$ is given by

$$\hat{U}_p(\delta\rho) = \hat{1} - \frac{i}{\hbar}(\delta\rho \cdot \hat{\mathbf{p}}). \tag{6.5^2}$$

A finite displacement ρ could be thought of as a succession of infinitesimal displacements. We obtain, then [analogously to Eq. (5.135)],

$$\hat{U}_p(\rho) = \exp\,[-(i/\hbar)\rho \cdot \hat{\mathbf{p}}]. \tag{6.5^3}$$

Problem 6.1: A symmetry transformation could also be defined in terms of its effect on the operators corresponding to the position, the linear momentum and the spin of the system. Thus $\hat{U}_\alpha(\delta\rho)$ is defined by [see Eq. (5.129)][9],

$$\hat{\mathbf{r}}' \equiv \hat{U}_\alpha(\delta\rho)\hat{\mathbf{r}}\hat{U}_\alpha^\dagger(\delta\rho) = \hat{\mathbf{r}} - (\delta\rho)\hat{1}, \tag{6.8a}$$

$$\hat{\mathbf{p}}' = \hat{U}_\alpha(\delta\rho)\hat{\mathbf{p}}\hat{U}_\alpha^\dagger(\delta\rho) = \hat{\mathbf{p}}, \tag{6.8b}$$

$$\hat{\mathbf{s}}' = \hat{U}_\alpha(\delta\rho)\hat{\mathbf{s}}\hat{U}_\alpha^\dagger(\delta\rho) = \hat{\mathbf{s}}, \tag{6.8c}$$

Show that this definition also leads to the result (6.7).

Thus, the generator of translations in space is the linear momentum operator. Invariance of the Hamiltonian under such translations requires [see, Eqs. (6.3a) and (6.1)], that the operator $\hat{U}_p$, and hence the operator $\hat{\mathbf{p}}$, commute with $\hat{H}$. But the relation $[\hat{\mathbf{p}}, \hat{H}] = \hat{0}$ implies that the total linear momentum of the system is conserved. In other words, *conservation of the linear momentum of a physical system is a consequence of the translational invariance of the Hamiltonian of the system.*

Now, the relation, $[\hat{\mathbf{p}}, \hat{H}] = \hat{0}$, will hold good only if $\hat{H}$ is independent of $\mathbf{r}$. That is, if $\hat{H}$ does not contain a potential energy term $\hat{V}$ which is a function of $\mathbf{r}$. In other words, the system should be *free* from external forces (or, *closed*). The degeneracy referred to in the previous Section, corresponds to the fact that the energy eigenvalues depend only on the magnitude of p and not on its direction.

The translational invariance of the Hamiltonian could also be expressed by saying that the transformed wavefunction $\psi'(\mathbf{r}, t) \equiv U_p(\rho)\psi(\mathbf{r}, t)$, satisfies the same Schrödinger equation as $\psi(\mathbf{r}, t)$. For, applying $\hat{U}_p(\rho)$ to the Schrödinger equation,

---

9. Note that the matrix elements of $\hat{\mathbf{r}}'$ with respect to the displaced states are the same as the matrix elements of $\hat{\mathbf{r}}$ with respect to the original states:

$$< \psi'_1(\mathbf{r}) \mid \hat{\mathbf{r}}' \mid \psi'_2(\mathbf{r}) > = < \psi_1(\mathbf{r}) \mid \hat{U}_\alpha^\dagger \hat{\mathbf{r}}' \hat{U}_\alpha \mid \psi_2(\mathbf{r}) >$$

$$= < \psi_1(\mathbf{r}) \mid \hat{\mathbf{r}} \mid \psi_2(\mathbf{r}) > .$$

$$i\hbar\frac{\partial\psi(\mathbf{r},t)}{\partial t}=\hat{H}\psi(\mathbf{r},t)$$

we have,

$$\begin{aligned} i\hbar\frac{\partial}{\partial t}\psi'(r,t)&=\hat{U}_p\hat{H}\psi(\mathbf{r},t)\\ &=\hat{H}\hat{U}_p\psi(\mathbf{r},t)\\ &=\hat{H}\psi'(\mathbf{r},t), \end{aligned} \tag{6.9}$$

since $$\hat{U}_p\hat{H}=\hat{H}\hat{U}_p.$$

6.2B Displacement in Time: Conservation of Energy

Displacement in time is also a continuous transformation and is, therefore, unitary. The operator corresponding to the infinitesimal displacement $\delta\tau$ in time of the physical system is given, according to Eq. (4.11), by[10]

$$\hat{U}_H(\delta\tau)=\hat{1}+(i/\hbar)\delta\tau\hat{H}, \tag{6.10^1}$$

and for a finite displacement τ, *when $\hat{H}$ is independent of time* [see Eq. (4.16)], by

$$\hat{U}_H(\tau)=\exp[i/\hbar)\tau\hat{H}], \tag{6.10^2}$$

$\hat{H}$ being the Hamiltonian of the system.

The effect of the displacement τ on the wavefunction $\psi(t)$ of the system, is given by [of. Eq. (6.6^1)],

$$\hat{U}_H(\tau)\psi(t)\equiv\psi'(t)=\psi(t-\tau). \tag{6.11}$$

This means that events corresponding to time t in ψ, correspond to time $(t+\tau)$ in ψ'.

The invariance of the Hamiltonian under translations in time requires that $\hat{U}_H(\tau)\hat{H}=\hat{H}\hat{U}_H(\tau)$, which also results in the same Schrödinger equation for $\psi'(t)$ as for $\psi(t)$ [see Eq. (6.9)]. Now, $\hat{H}$ commutes with $\hat{U}_H$ if the latter is given by expression (6.10^2); that is, if $\hat{H}$ is independent of t. And the time-independence of $\hat{H}$ means that the total energy of the physical system is conserved. But when $\hat{H}$ depends on t (that is, when the energy is not conserved), $\hat{U}_H(\tau)$ is not given by Eq. (6.10^2), but by a more complicated expression [see Eq. (4.13)] which does not commute with $\hat{H}$. As a result, the Hamiltonian would not be invariant under translation in time[11]. We, thus, see that *the total energy of the system is conserved if the system is invariant under translations in time, and vice versa.*

---

10. In place of the negative sign in Eq. (4.11), we have the positive sign in (6.10^1), because a displacement $\delta\tau$ in time of the physical system corresponds to a displacement $-\delta\tau$ for the time co-ordinate [c.f. Eq. (6.6^1)].

11. This fact also follows from the fact that $\psi'(t)=\hat{U}_H(\delta\tau)\psi(t)$ does not satisfy the Schrödinger equation when $\hat{H}$ in Eq. (6.10^1) is a function of time (see Problem 6.2).

Problem 6.2: Show that $\psi'(t) = \hat{U}_H(\delta\tau)\psi(t)$, will not satisfy the Schrödinger equation when $\hat{H}$ is a function of time.

6.2C Rotations in Space: Conservation of Angular Momentum

We have already discussed, in Section 5.6, the relationship of rotations to the angular momentum of the physical system. The unitary operator corresponding to a rotation of the physical system through an angle θ about the axis $\mathbf{n}$, is given by Eq. (5.135):

$$\hat{U}_J(\mathbf{n}, \theta) \equiv \hat{R}_n(\theta) = \exp\left[-(i/\hbar)\theta(\mathbf{n}.\hat{\mathbf{J}})\right] \tag{6.12}$$

where, $\hat{\mathbf{J}}$ is the operator corresponding to the *total* angular momentum of the system. Hence the invariance of the Hamiltonian under rotations in space ($\hat{U}_J\hat{H} = \hat{H}\hat{U}_J$), requires that the total angular momentum be a constant of motion (that is, $[\hat{\mathbf{J}}, \hat{H}] = \hat{0}$). In other words, *conservation of angular momentum is a consequence of the rotational invariance of the system.*

Problem 6.3: Show that the energy eigenfunctions of a spinless system with a spherically symmetric potential, are given by $\psi_E(\mathbf{r}) = R_E(r)Y_{lm}(\theta, \phi)$, where, Y_{lm} is the spherical harmonic of order l and R_E is a function only of the radial co-ordinate.

6.2D Space Inversion: Parity

Space inversion is the operation in which the axes of the reference frame are reflected at the origin, as in Fig. 5.2. The effect of the operation is to change a right-handed co-ordinate system into a left-handed one[12], and *vice versa*. Since these two types of co-ordinate systems are mirror images of each other, a physical system which is invariant under space inversion is said to possess *reflection symmetry*, or to lack *chirality* (that is, handedness). This means that phenomena associated with the system will look the same irrespective of the handedness of the co-ordinate system in which they are observed. In other words, there is parity between the system and its mirror image. Space inversion is, for this reason, also called *Parity Operation*. The corresponding transformation is, obviously, discrete.

Now, an element of the physical system which has the position vector $\mathbf{r}$ in the original co-ordinate system, has the position vector $\mathbf{r}' = -\mathbf{r}$ in the inverted co-ordinate system (see Fig. 5.2). The velocity $\frac{d\mathbf{r}}{dt}$, and hence the momentum $\mathbf{p}$, also

12. A right-handed co-ordinate system is one in which, if we twist a right handed screw from x to y, it will advance in the direction of the z-axis. A left-handed co-ordinate system is similarly defined by a left-handed screw.

change sign under the operation. But the angular momentum $\mathbf{r} \times \mathbf{p}$ will not change sign. Thus, the parity operator $\hat{U}_P$ could be defined by the relations[13] [see Eqs. (6.8a-c)].

$$\hat{\mathbf{r}}' \equiv \hat{U}_P^\dagger \hat{\mathbf{r}} \hat{U}_P = -\hat{\mathbf{r}}, \tag{6.13a}$$

$$\hat{\mathbf{p}}' \equiv \hat{U}_P^\dagger \hat{\mathbf{p}} \hat{U}_P = -\hat{\mathbf{p}}, \tag{6.13b}$$

$$\hat{\mathbf{s}}' \equiv \hat{U}_P^\dagger \hat{\mathbf{s}} \hat{U}_P = \hat{\mathbf{s}}, \tag{6.13c}$$

We see that the transformation preserves the basic Heisenberg commutation relations (3.12a, b) and is, therefore, unitary (rather than antiunitary). Also, multiplying Eqs. (6.13a-c) on the right by $\hat{U}_P^\dagger$ and on the left by $\hat{U}_P$, we see that

$$\hat{U}_P^\dagger = \exp(i\delta)\hat{U}_P, \tag{6.14a}$$

which, in view of the relation, $\hat{U}_P^\dagger \hat{U}_P = \hat{1}$, implies,

$$\hat{U}_P^2 = \exp(-i\delta)\hat{1}. \tag{6.14b}$$

Now, at least in the case of systems with integer spin [see Eq. (5.145c)], $\hat{U}_P^2 = \hat{1}$, so that,

$$\delta = 0. \tag{6.15}$$

Thus, $\hat{U}_P$ is both unitary and *Hermitian*. Therefore we can set,

$$\hat{U}_P \equiv \hat{P}. \tag{6.16}$$

Then, from Eqs. (6.14b) and (6.15), we see that the eigenvalues of $\hat{P}$ are ± 1 (Remember that the eigenvalues have to be real since $\hat{P}$ is Hermitian). Thus, if $\psi(\mathbf{r})$ is an eigenvector of $\hat{P}$, then [by Eqs. (6.6[1]) and (6.13a)],

$$\hat{P}^\dagger \psi(\mathbf{r}) \equiv \hat{P}\psi(\mathbf{r}) = \psi(-r) = \pm\psi(\mathbf{r}). \tag{6.17}$$

A state for which the upper sign in Eq. (6.17) holds good, is said to have *even*, or *positive*, parity whereas one for which the lower sign is applicable is of *odd*, or *negative*, parity. Of course, a physical state need to have a definite parity (that is, need be an eigenvector of $\hat{P}$) only if the Hamiltonian (interaction) is invariant under the parity operation. So far only the weak interaction (footnote 7) is known to violate parity conservation. This means that phenomena mediated by weak interaction, such as beta decay, can be distinguished from their mirror images[14].

An operator also could be characterised as odd or even depending on whether or not it changes sign under the parity operation. The operator $\hat{A}$ is *odd*, if,

---

13. Relation (6.13c) is based on the assumption that the spin transforms in the same way as the orbital angular momentum under space inversion [cf. Eqs. (5.128[2]) and (5.127[3])].
14. For a more detailed discussion of the experimental aspects of parity violation, see A. Bohr and B. R. Mottelson, *Nuclear Structure*, Vol. I (Benjamin, New York 1969), Section 1-2.

$$\hat{P}\hat{A}\hat{P}^{\dagger} = -\hat{A}, \tag{6.18a}$$

or,

$$\hat{A}\hat{P} + \hat{P}\hat{A} \equiv \{\hat{A}, \hat{P}\} = \hat{0}; \tag{6.18b}$$

and *even* if,

$$\hat{P}\hat{A}\hat{P}_{\dagger} = \hat{A}, \tag{6.19a}$$

or,

$$[\hat{A}, \hat{P}] = \hat{0}. \tag{6.19b}$$

Thus, $\hat{A}$ is even or odd according as it commutes or anticommutes with $\hat{P}$. Obviously, then, only the eigenstate of an even operator can have definite parity. For example, Eq. (6.13b) shows that $\hat{\mathbf{p}}$ is an odd operator. Therefore, a state with definite linear momentum has no definite parity. On the other hand, the angular momentum operator is even [Eq. (6.13c)], so that a state with a definite angular momentum has a definite parity. In fact, we have shown in Section 5.4 [Eq. (5.61)] that a state with orbital angular momentum quantum number l, has the parity $\pi_l = (-1)^l$. Eq. (6.13c) shows that the spin part of the wavefunction (χ_s) also has a definite parity, usually referred to as the *intrinsic parity* of the particle (spin being the *intrinsic* angular momentum). However, the intrinsic parity is not given by a simple formula involving the spin quantum number s, but has to be determined relative to a particle whose intrinsic parity is already fixed either by convention or by some other means. For example, the intrinsic parity of the nucleons (spin $\frac{1}{2}$) is conventionally fixed as positive. Then, the parity of the pions can be determined relative to that of the nucleons, for example[15], from the reaction,

$$\pi^{+} + d \rightarrow p + p,$$

where d is the deuteron (consisting of a proton and a neutron), p represents the proton and π^{+} a positively charged *pion*. The parity of the *pion* turns out to be negative.

Problem 6.4: The parity π_A of an operator $\hat{A}$ is defined by $\hat{P}\hat{A}\hat{P}^{\dagger} = \pi_A\hat{A}$; and the parity of a state vector $|\psi>$ by $\hat{P}\,|\psi> = \pi_\psi\,|\psi>$. Show that the selection rule, $\pi_\phi\pi_A\pi_\psi = +1$, applies to the matrix element $<\phi\,|\,\hat{A}\,|\,\psi>$ of $\hat{A}$. Hence deduce that a nucleon cannot have an electric dipole moment.

(*Note*: Electric dipole moment operator, $\hat{\mathbf{D}} = e\hat{\mathbf{r}}$, e being the electric charge of the nucleon).

---

15. See, for example, Omnes, R., *Introduction to Particle Physics* (Wiley-Interscience, London 1971; translated by G. Barton) Chapter 6, Section 4; A.D. Martin and T.D. Spearman, *Elementary Particle Theory* (North-Holland, Amsterdam 1970), Chapter 5, Section 2.1.

6.2E Time Reversal Invariance

In time reversal operation $t \to -t$. Under such an operation, the velocity changes sign, whereas the position co-ordinate is not affected. Thus, denoting the time reversal operator by $\hat{\mathcal{T}}$, we have,

$$\bar{\hat{\mathbf{r}}} = \hat{\mathcal{T}}\hat{\mathbf{r}}\hat{\mathcal{T}}^{\dagger} = \hat{\mathbf{r}}, \tag{6.20a}$$

$$\bar{\hat{\mathbf{p}}} = \hat{\mathcal{T}}\hat{\mathbf{p}}\hat{\mathcal{T}}^{\dagger} = -\hat{\mathbf{p}}, \tag{6.20b}$$

$$\bar{\hat{\mathbf{s}}} = \hat{\mathcal{T}}\hat{\mathbf{s}}\hat{\mathcal{T}}^{\dagger} = -\hat{\mathbf{s}}, \tag{6.20c}$$

This transformation leads to the commutation relations (B.14) and is, hence, *antiunitary*. It follows that, unlike parity, time reversal invariance does not give rise to an associated observable[16]. Therefore, in the equation,

$$\hat{\mathcal{T}}^2 = e^{i\gamma}\hat{1}, \tag{6.21}$$

which is derived analogously to Eq. (6.14b), γ need not be zero.

The time-reversed state $\psi_{rev}(\mathbf{r}, t)$ corresponding to the state $\psi(\mathbf{r}, t)$ is not $\psi(\mathbf{r}, -t)$ [cf. Eq. (6.17)]. This could be seen as follows: Let $\hat{H}$ be the (real) Hamiltonian of a physical system that is invariant under time reversal. That is,

$$\hat{\mathcal{T}}\hat{H}\hat{\mathcal{T}}^{\dagger} = \hat{H}, \text{ or, } [\hat{\mathcal{T}}, \hat{H}] = \hat{0}. \tag{6.22}$$

The Schrödinger equation for the original state $\psi(\mathbf{r}, t)$ is,

$$i\hbar\frac{\partial\psi(\mathbf{r}, t)}{\partial t} = \hat{H}\psi(\mathbf{r}, t). \tag{6.23a}$$

Replacing t by $-t$ in this equation, we have,

$$-i\hbar\frac{\partial\psi(\mathbf{r}, -t)}{\partial t} = \hat{H}\psi(\mathbf{r}, -t). \tag{6.23b}$$

On the other hand, operating on Eq. (6.23a) with $\hat{\mathcal{T}}$ and making use of the relations (6.22) and (B.2), we get,

$$i\hbar\frac{\partial}{\partial t}\{\hat{\mathcal{T}}\psi(\mathbf{r}, t)\} = \hat{H}\{\hat{\mathcal{T}}\psi(\mathbf{r}, t)\}. \tag{6.23c}$$

Comparison of Eqs. (6.23b) and (6.23c) shows that $\hat{\mathcal{T}}\psi(\mathbf{r}, t) \neq \psi(\mathbf{r}, -t)$. However, taking the complex conjugate of (6.23b), we get,

$$i\hbar\frac{\partial}{\partial t}\psi^*(\mathbf{r}, -t) = \hat{H}\psi^*(\mathbf{r}, -t), \tag{6.23d}$$

which, in view of Eq. (6.23c) suggests,

$$\hat{\mathcal{T}}\psi(\mathbf{r}, t) \equiv \psi_{rev}(\mathbf{r}, t) = \psi^*(\mathbf{r}, -t). \tag{6.24a}$$

This result could be obtained also from the following considerations:

---

16. It is in anticipation of this result that we have denoted the time reversal operator by $\hat{\mathcal{T}}$ rather than by $\hat{U}_{\mathcal{T}}$

$\hat{T}$, being an antiunitary operator, could be written as the product of a unitary operator $\hat{U}$ and the complex conjugation operator $\hat{K}_c$ [See Eq. (B.20a)]:

$$\hat{T} = \hat{U}\hat{K}_c. \tag{6.25}$$

In the co-ordinate representation, $\hat{\mathbf{r}}$ is real and $\hat{p}$ is pure imaginary [Eqs. (3.18[1])]. Therefore, in the case of a *spinless* system, Eqs. (6.20a, b) suggest the choice,

$$\hat{U} = \hat{1};\ \hat{T} = \hat{K}_c. \tag{6.25[1]}$$

Thus, $\hat{T}$ is the complex conjugation operator. This fact, of course, depends on the representation. In the momentum representation, $\hat{p}$ is real and $\hat{\mathbf{r}}$ is imaginary [Eqs. (3.18[2])]; so that Eqs. (6.20a, b) require $\hat{U}$ to be an operator that changes $\hat{\mathbf{p}}$ to $-\hat{\mathbf{p}}$: $\hat{U}\hat{\mathbf{p}}U^{+} = -\hat{\mathbf{p}}$, and

$$\hat{T}\psi(\mathbf{p}, t) = \psi^{*}(-\mathbf{p}, -t), \tag{6.24b}$$

$\hat{T}$ for Particles with Spin

Now, it is clear from the foregoing that the explicit form for $\hat{T}$ will depend on the representation chosen. We, therefore, choose the co-ordinate representation for $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$ [Eqs. (3.18[1])] and the standard representation for the angular momentum [Eq. (5.32a-d)]. Then, $\hat{\mathbf{r}}$, $\hat{s}_x$ and $\hat{s}_z$ are real whereas $\hat{\mathbf{p}}$ and $\hat{s}_y$ are imaginary. If $\hat{T}$ is written as in (6.25), then from Eqs. (6.20a-c), we have,

$$\hat{U}\hat{\mathbf{r}}\hat{U}^{\dagger} = \hat{\mathbf{r}};\ \hat{U}\hat{\mathbf{p}}\hat{U}^{\dagger} = \hat{\mathbf{p}};\ \hat{U}\hat{s}_y\hat{U}^{\dagger} = \hat{s}_y. \tag{6.26a}$$

$$\hat{U}\hat{s}_x\hat{U}^{\dagger} = -\hat{s}_x;\ \hat{U}\hat{s}_z\hat{U}^{\dagger} = -\hat{s}_z. \tag{6.26b}$$

$\hat{U}$ leaves $\hat{\mathbf{r}}$, $\hat{\mathbf{p}}$ and $\hat{s}_y$ invariant, but changes the signs of $\hat{s}_x$ and $\hat{s}_z$. $\hat{U}$ should, therefore, represent a rotation of the co-ordinate system through an angle π about the y-axis in the spin space. That is[17] [see Eq. (5.135)],

$$\hat{U} = e^{(i/\hbar)\pi\hat{s}_y}. \tag{6.27}$$

Thus, in the case of a system with total angular momentum $\hat{\mathbf{J}}$,

$$\hat{T} = \exp\left[(i/\hbar)\pi\hat{J}_y\right]\hat{K}_c, \tag{6.25[2]}$$

so that, if $|\, jm >$ represents a basis vector in the angular momentum representation, then,

$$\hat{T}|\, jm > = (-1)^{j+m}\,|\, j-m >. \tag{6.28}$$

The result could be derived as follows:

$$\begin{aligned}\hat{T}|\, jm > &= \exp\left[(i/\hbar)\pi\hat{J}_y\right]\cdot\hat{K}_c\,|\, jm > \\ &= \exp\left[(i/\hbar)\pi\hat{J}_y\right]\,|\, jm >, \text{ by Eq. (B.17).} \\ &= \sum_{m'} d^{j}_{m'm}(-\pi)\,|\, jm' >, \end{aligned} \tag{6.28a}$$

---

17. The exponent in Eq. (6.27) has positive sign because $\hat{U}$ represents a rotation of the co-ordinate system rather than of the physical system.

where, Eqs. (5.136[2]), (5.138[2]) and (5.141) have been used. But, from Eq. (5.142), we have,

$$d^{j}_{m'm}(-\pi) = (-1)^{j+m}\delta_{m',-m} \tag{6.29}$$

Substitution of (6.29) in (6.28a) leads to (6.28).

Applying $\hat{T}$ once again to Eq. (6.28), we get,

$$\hat{T}^2 | jm > = (-1)^{2j} | jm > . \tag{6.30}$$

Thus, $\hat{T}^2$ has eigenvalue +1 for integer j and − 1 for half-integer j. If the Hamiltonian $\hat{H}$ is time-reversal invariant, then $\hat{T}| jm >$ would be an eigenvector[18] of $\hat{H}$ belonging to the same eigenvalue as $| jm >$. The case $\hat{T}^2 | jm > = | jm >$, then, permits the solution, $\hat{T}| jm > = c | jm >$, where c is a scalar. In this case, $\hat{T}| jm >$ is the same state as $| jm >$ so that there is no degeneracy. But, when $\hat{T}^2 | jm > = - | jm >, \hat{T}| jm > \neq c | jm >$. In fact, $\hat{T}| jm >$ would be orthogonal to $| jm >$. For,

$$\begin{aligned} < jm | \{\hat{T}| jm >\} &= < jm | \{\hat{T}^{\dagger}\hat{T}^2 | jm >\} \\ &= - < jm | \{\hat{T}^{\dagger} | jm >\} \\ &= - < jm | \{\hat{T}| jm >\}\text{; by Eq. (B.6b).} \end{aligned}$$

This shows that $< jm | \{\hat{T}| jm >\}$ is zero. Thus, every state $| jm >$ would be degenerate with its time-reversed state. In other words, every energy eigenvalue would be two-fold degenerate. This is referred to as *Kramer's degeneracy.*

An example of a state for which $\hat{T}^2 = \hat{1}$, is the state $| lm >$ of a system with orbital angular momentum l. For, according to Eq. (6.30),

$$\hat{T}^2 | lm > = (-1)^{2l} | lm > = lm >, \tag{6.30a}$$

since l is integer. A state of an atom with an odd number of electrons (hence, half-odd j), provides an example of Kramer's degeneracy. The degeneracy would be present even in the presence of an electric field (as the potential $\hat{V}(\mathbf{r}) = -\hat{\mathbf{E}} \cdot \hat{\mathbf{r}}$, commutes with $\hat{T}$), but would be lifted by a velocity-dependent field such as the magnetic field.

When the angular momentum is entirely due to orbital motion ($j = l$), we can take the co-ordinate representation[19] of Eq. (6.28). We get,

$$< \mathbf{r} | \{\hat{T}| lm >\} = (-1)^{l+m} < \mathbf{r} | l - m > .$$

But, using Eq. (B.6b), we have,

$$< \mathbf{r} | \{\hat{T}| lm >\} = \{< \mathbf{r} | \hat{T}\} | lm >^{*} = < \mathbf{r} | lm >^{*},$$

since,

$$\hat{T}| \mathbf{r} > = \hat{K}_c | \mathbf{r} > = | \mathbf{r} >, \tag{6.31}$$

---

18. The basis vector $| jm >$ is actually labelled by an additional quantum number α corresponding to the eigenvalues of $\hat{H}$. Thus,

$$| jm > \equiv | \alpha jm > .$$

19. Note that the intrinsic spin has no co-ordinate representation.

in the co-ordinate representation. Thus,

$$< \mathbf{r} \mid lm >^{*} = (-1)^{l+m} < \mathbf{r} \mid l-m > . \qquad (6.28^1)$$

The left hand side of Eq. (6.28[1]) is the co-ordinate representation of $\hat{T} \mid lm >$. We see, therefore, that, in the co-ordinate representation, time reversal is equivalent to complex conjugation except for systems with intrinsic spin. Also, comparing Eq. (6.28[1]) with Eqs. (5.46) and (5.59a), we see that the choice of phase implied by Eq. (6.28) is given by

$$< \mathbf{r} \mid lm > = \mathcal{Y}_{lm}(\theta, \phi) = i^l Y_{lm}(\theta, \phi), \qquad (6.32)$$

for the eigenvectors of the orbital angular momentum[20]. With this choice, complex conjugation becomes identical with rotation through π about the y-axis, while either of the operations becomes equivalent to the time reversal operation[21].

It might seem strange that a *linear* operator corresponding to a rotation could be equivalent to $\hat{T}$ which is an *antilinear* operator. But, as is evident from Eqs. (B. 1, 2), linear and antilinear operations are different only with regard to 'complex' entities. The relation

$$\hat{K}_c \mid lm > = \mid lm >,$$

which is implicit in Eq. (6.28[1]), suggests that the $\mathcal{Y}_{lm}(\theta, \phi)$'s of Eq. (6.32) constitute a 'real' basis for the orbital angular momentum.

Problem 6.5: Show from Eqs. (6.27) and (6.25) that, in the case of a spin $\frac{1}{2}$ particle, $\hat{T} = i\hat{\sigma}_y \hat{K}_c$, where $\hat{\sigma}_y$ is the Pauli spin operator.

Transformation Properties of Spherical Tensors

The spherical tensors have usually simple transformation properties under time reversal. In analogy with that of the spherical harmonic, the transform $\overline{\hat{T}}_q^{(k)}$ of a spherical tensor $\hat{T}_q^{(k)}$ of rank k under time reversal, can be expressed as

$$\overline{\hat{T}}_q^{(k)} \equiv \hat{T} \hat{T}_q^{(k)} \hat{T}^{\dagger} = \pm (-1)^q \hat{T}_{-q}^{(k)}. \qquad (6.33)$$

The spherical tensor is said to be *even* if the upper sign in (6.33) holds good, and *odd* if the lower sign holds good. For example, the spherical components of $\hat{\mathbf{r}}$ are given by [see Eq. (5.167)],

$$\hat{r}_{\pm 1}^{(1)} = \pm \frac{1}{\sqrt{2}} (\hat{x} \pm i\hat{y}); \; \hat{r}_0^{(1)} = \hat{z}, \qquad (6.34)$$

---

20. As we have explained in relation to Eqs. (5.29a, b) and (5.57), we are free to choose the phase of one of the eigenvectors belonging to a given j- or l-value.
21. In the present case, the time reversal operation is just complex conjugation. It is the particular choice of phase, Eq. (6.32), that makes this operation also equal to a rotation through π about the y-axis.

so that, using Eq. (6.20a), we get,

$$\hat{T}\hat{r}_q^{(1)}\hat{T}^\dagger = +(-1)^q \hat{r}_{-q}^{(1)}. \tag{6.33a}$$

Thus, $\hat{r}_q^{(1)}$ is even, or $\hat{\mathbf{r}}$ is an even operator. Similarly, $\hat{\mathbf{p}}$ is an odd operator. It follows that the electric dipole moment operator $\hat{D} = e\hat{\mathbf{r}}$ (where, e is the electric charge) is even under time reversal. Now, it can be shown, using Eq. (6.33), that

$$< j||\hat{T}^{(k)}||j > = \pm(-1)^k < j||\hat{T}^{(k)}||j >^*, \tag{6.35}$$

where, the double-bar denotes reduced matrix element between states of definite angular momenta [Eq. (5.170[2])]. Also, the sign in Eq. (6.35) follows that in Eq. (6.33). Eq. (6.35) shows that a stationary state with a definite angular momentum cannot have an electric dipole moment, unless time-reversal invariance is violated. Earlier [Problem 6.4], we have seen that parity conservation forbids the existence of an electric dipole moment for the nucleon. Therefore, the observation of a static electric dipole moment for the nucleon would imply the violation of both parity conservation and time-reversal symmetry.

Problem 6.6: Deduce Eq. (6.35).

CHAPTER 7

THEORY OF SCATTERING

7.1 PRELIMINARIES

Experiments on scattering provide one of the most important means of gathering information in the realm of atomic and subatomic particles. Thus, it was Rutherford's experiments on the scattering of alpha-particles that provided the experimental basis for the nuclear atom model that eventually led to the Bohr atom model and to quantum mechanics. Many of the features of the nuclear force such as its range, strength and spin-dependence, have been deduced from data gathered from nucleon-nucleon scattering. Scattering of electrons from nuclei as well as nucleons has helped in determining the charge distribution in the latter. In fact, the importance of the role of the scattering experiment as a peep-hole into the world of nuclear and sub-nuclear particles, cannot be overemphasized.

In this chapter, we propose to present some aspects[1] of the quantum theory of scattering. After describing the scattering experiment and the quantities measured in the experiment, we will consider two different methods of the theory, one suitable for low energy scattering and the other valid for high energy scattering.

The Scattering Experiment

In a typical scattering experiment, a *beam* of homogeneous, monoergic particles (that is, an *ensemble* of particles) falls on a *target* (which could be in the form of a thin foil) consisting of a large number of *scattering centres*. The particles are scattered by the target in all directions[2] and the scattered particles are received and analysed by a *detector* placed at a large (compared with the linear dimensions of the target) distance from the target. Let (see Fig. 7.1) the origin of the co-ordinate system be chosen at the target, and the z-axis along the direction of the incident beam. Also let,

---

1. Sufficient as it is to be the subject matter of a whole book [See, for example, Goldberger, M.L. and Watson, K.M. *Collision Theory* (John Wiley, New York 1964); Newton, R.G., *Scattering Theory of Waves and Particles* (McGraw-Hill, New York 1966)].
2. Some of the particles, of course, proceed along the incident direction, unaffected by the target.

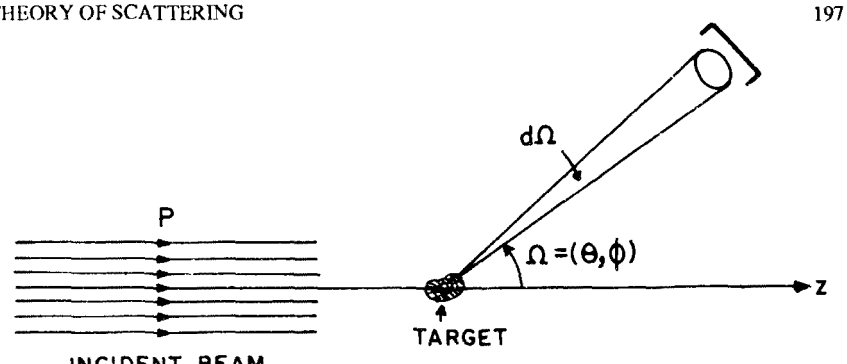


Fig. 7.1 The scattering experiment.

$d\Omega \equiv \sin\theta d\theta d\phi$ = the solid angle subtended by the detector at the target,

$\Omega \equiv (\theta, \phi)$ = the angular position of the detector,

dN = the number of particles received by the detector per unit time,

$\mathbf{J}_0$ = the incident flux,

= the number of particles crossing a unit area perpendicular to the direction of the beam, per unit time,

n = the number of scattering centres in the target.

Then, dN will be proportional to n, $|\mathbf{J}_0|$ and $d\Omega$, if the following conditions are satisfied:

(i) The intensity (that is, the density or the number of particles per unit volume) of the incident beam is low enough for the mutual interaction among the particles to be neglected.

(ii) The momentum p of the particles is large enough (so that the de Bröglie wavelength h/p is small enough) for the scattering centres to act independently of each other in scattering the particles.[3] Satisfying these conditions is not an entirely 'painless' affair. For, condition (i) conflicts with the requirement for 'good statistics' which favours a large intensity (the larger the number of 'counts' at the detector during a given time, the better the statistics). Similarly, if p becomes too large, then phenomena more complicated than simple scattering, such as particle production, will begin to appear. Thus, there are optimum values for J_0 and p for a given set up.

When the conditions are satisfied, we have,

$$dN = \sigma(\Omega) n \cdot |\mathbf{J}_0| \, d\Omega. \qquad (7.1^1)$$

It is easy to see that the proportionality factor $\sigma(\Omega)$ has the dimension of an area. Let $\mathbf{J}_s(\Omega)$ represent the scattered flux of particles in the direction Ω. Then,

$$dN = |\mathbf{J}_s| \, r^2 d\Omega, \qquad (7.1^2)$$

where, r is the distance of the detector from the target. From (7.1^1) and (7.1^2), we get,

3. The actual condition is that the distance between the scattering centres in the target be large compared with the de Bröglie wave length of the incident particles.

$$\sigma(\Omega) = \frac{r^2 |\mathbf{J}_s|}{n |\mathbf{J}_0|}, \tag{7.2}$$

which proves our earlier statement. Eq. (7.1[1]) shows that $\sigma(\Omega)$ is numerically equal to the cross-sectional area of the incident beam that is traversed by as many particles as are scattered per scattering centre, into a unit solid angle in the direction $\Omega \cdot \sigma(\Omega)$ is, for this reason, called the *differential scattering cross section*. The integral of $\sigma(\Omega)$ over the solid angles, is the *total scattering cross section* which, according to the above interpretation, is the cross-sectional area of the incident beam that is traversed by as many particles as are scattered in all directions by each scattering centre. Thus[4],

$$\sigma_{tot} = \int \sigma(\Omega) d\Omega. \tag{7.3}$$

The differential and the total scattering cross sections could be regarded as the experimental quantities of the scattering problem. The aim of the scattering theory is, then, to make use of these to deduce information on the force, or the interaction, responsible for the scattering. The theory achieves this aim by establishing a relationship between the cross-section and the wavefunction of the system. For, the wavefunction is obtained by solving the Schrödinger equation which involves the interaction. Thus, whereas in a bound state problem the emphasis is on the energy eigenvalues, in the scattering problem the interest centres on the wavefunctions. Also, in the bound state problem one is dealing with the negative part of the energy spectrum, while in the scattering problem it is the positive part of the spectrum that comes into play.

Before deriving a relationship between the cross-section and the wave function, let us specify the type of scattering problems to which we will confine our discussion. Scattering is said to be *elastic* if there is only momentum transfer, but no energy transfer, taking place between the incident particles and the target.[5] In this case, the particles undergo only a change in their direction of motion but suffer no change in the magnitude of their momentum. On the other hand, if there is exchange of energy between the incident particles and the target, the scattering is *inelastic*. We will consider only the first type of scattering.

Also, as is implied by the choice of the origin at the target, we are assuming that the target is fixed. Elastic scattering by such a target can be described as *scattering by a potential* $V(r)$.

---

4. The differential scattering cross-section is sometimes denoted by the symbol $\left(\frac{d\sigma}{d\Omega}\right)$. The advantage of this notation is that the symbol for the total scattering cross-section does not require a subscript; for,

$$\sigma_{tot} \equiv \int\left(\frac{d\sigma}{d\Omega}\right) d\Omega = \int d\sigma = \sigma.$$

We will hereafter adopt this notation.

5. This definition applies to *potential scattering* (that is, scattering by a potential). In other types of scattering such as scattering of nucleons by nuclei, the scattering is elastic even if transfer of (kinetic) energy takes place without producing any internal excitation.

Now, in an actual experiment, the scattering centres suffer recoil. The co-ordinate system, in which this recoil is explicitly taken into account, is known as the *Laboratory System*[6] (Fig. 7.2). However, when the interaction depends only on the relative co-ordinate $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, the Schrödinger equation corresponding to

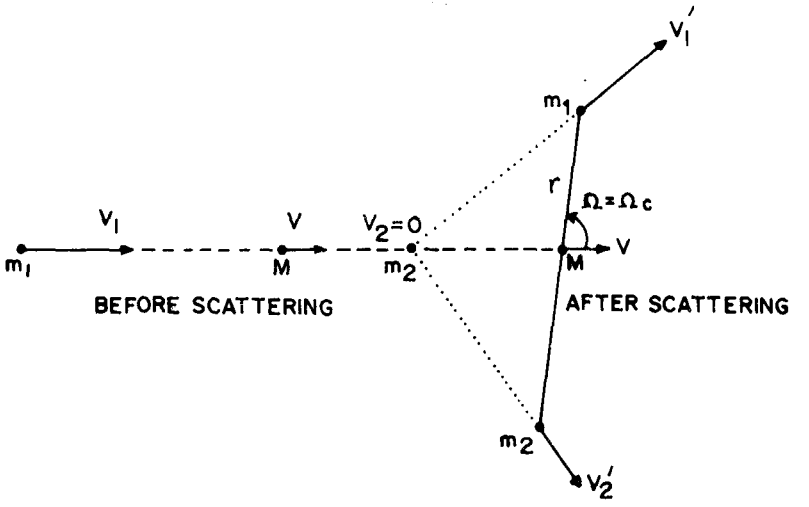


Fig. 7.2 Two-particle scattering in the Laboratory System of coordinates. The Centre-of-Mass velocity $\mathbf{V} = (m_1/M)\mathbf{v}_1$, where $M = m_1 + m_2$, $\mathbf{v}'_2$ is the recoil velocity.

the scattering of one particle by another, could be separated into two parts: one corresponding to a uniform motion of the centre-of-mass of the system and the other corresponding to the relative motion of the particles [see, Section 4.2B]. The relative motion *appears* as the motion of a single particle of mass $\mu = \dfrac{m_1 m_2}{m_1 + m_2}$

(where m_1 and m_2 are, the masses of the colliding particles, and the μ is referred to as the *reduced mass*) with a velocity $\mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2$, under the influence of a potential $V(\mathbf{r})$. Scattering affects only the relative motion which is independent of the velocity of the centre-of-mass. We see, thus, that the theory of the scattering of a particle by a potential is also the theory of the two-particle scattering problem.

Relationship of the Scattering Cross-section to the Wave function: The Scattering Amplitude

It follows from the foregoing, that the (time-independent) Schrödinger equation of the scattering problem is given by

---

6. The Laboratory system is also often defined as the co-ordinate system in which the target is *initially* at rest, whereas, the co-ordinate system in which the centre-of-mass is at rest, is called the *Centre-of-Mass System*. For a detailed discussion of the relationship between the two systems, see, for example. A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam 1961), Chapter 10, Section 7.

$$\left[-\frac{\hbar^2}{2\mu}\Delta + V(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{7.4}$$

where, $\Delta \equiv \nabla^2$, the Laplacian, and

$$E \equiv E_k = \frac{\hbar^2 k^2}{2\mu}, \tag{7.5}$$

$\hbar\mathbf{k} = \mathbf{p}$, being the momentum of the incident particles. We will assume that the potential $V(\mathbf{r})$ is of finite range and that it goes to zero faster than $\frac{1}{r}$ as $r \to \infty$. This excludes the pure Coulomb force, but includes the *screened* Coulomb force such as that is experienced by electrons being scattered by atoms.

We seek a solution to Eq. (7.4) subject to the following conditions:

(i) Initially (long before the scattering), it represents a homogeneous beam of particles of momentum $\mathbf{p} = \hbar\mathbf{k}$, proceeding in the positive z-direction. That is[7]

$$\psi_{\text{initial}} \equiv \psi_k^i(\mathbf{r}) \underset{r\to\infty}{\sim} \exp(i\mathbf{k}\cdot\mathbf{r}) = \exp(ikz). \tag{7.6}$$

(ii) Finally (long after scattering), it represents particles issuing in all directions from the scattering centre. Such a flux of particles is represented by a *spherical wave* (in contrast to the expression (7.6) which corresponds to a *plane wave*[8]) whose intensity falls off inversely as the square of the distance from the centre. Thus,

$$\psi_{\text{final}} \equiv \psi_k^s(\mathbf{r}) \underset{r\to\infty}{\sim} f_k(\Omega)\frac{e^{ikr}}{r}. \tag{7.7}$$

Here, $f_k(\Omega)$ is independent of r, and is known as the *scattering amplitude*.

Thus, the general solution of Eq. (7.4) corresponding to the scattering problem can be written as,

$$\psi_k(\mathbf{r}) \underset{r\to\infty}{\sim} e^{ikz} + f_k(\Omega)\frac{e^{ikr}}{r}. \tag{7.8}$$

The first term on the R.H.S. in (7.8) represents the transmitted particles (particles that are unaffected by the potential) and the second term particles that are scattered.

---

7. We omit the normalization factor $(2\pi)^{3/2}$ because, in the scattering problems, we are interested in the relative probability (out of N particles incident in the z-direction, how many are scattered in the direction Ω ?) rather than in the absolute one. Unnormalized wavefunctions are, therefore, more convenient to work with.

8. A plane wave really corresponds to a uniform distribution of particles all over space, since the probability density $|e^{i\mathbf{k}\cdot\mathbf{r}}|^2$ is independent of r. The plane wave provides an approximate representation of the incident beam for the scattering problem when the transverse extension of the beam is large compared with the range of the potential, or linear dimensions of the scatterer. The latter is of the order of 10^{-8} cm for atoms and 10^{-12} cm for nuclei.

Density of particles in the incident beam $= \rho_{inc} = |\psi_k^i(\mathbf{r})|^2 = 1.$

Velocity of the particles $= v_i = \frac{\hbar k}{\mu}.$

Incident flux $= |J_0| = \rho_{inc} v_i = \frac{\hbar k}{\mu}.$

Density of scattered particles in the direction Ω $= \rho_{scat} = \frac{|f_k(\Omega)|^2}{r^2}$

Velocity of the scattered particles $= v_s = v_i = \frac{\hbar k}{\mu}.$

Scattered flux in the direction Ω $= |J_s(\Omega)|$

$= \rho_{scat} v_s$

$= \frac{\hbar k}{\mu} |f_k(\Omega)|^2 / r^2$

Thus, from Eq. (7.2), we get[9],

$$\frac{d\sigma}{d\Omega} \equiv \sigma(\Omega) = |f_k(\Omega)|^2 . \tag{7.9}$$

The method of solution would, thus, consist in finding a solution to Eq. (7.4) valid for large values of r and writing the solution in the form of a transmitted plane wave and an outgoing spherical wave. The amplitude of the spherical wave is then $f_k(\Omega)$. $\frac{d\sigma}{d\Omega}$ is obtained from Eqs. (7.9) and (7.3).

7.2 METHOD OF PARTIAL WAVES

This is a method which can be applied when the potential is central [$V(\mathbf{r}) \equiv V(r)$] and is of finite range [$V(r) \approx 0, r > r_0$]. We will see later that the method is really useful only when the energy of the incident particles is rather low. The method exploits the fact that the angular momentum of a particle is conserved in a central field. Therefore, if we categorise the particles according to their angular momenta (classically, this amounts to categorising the particles according to their *impact parameters*[10], since all the particles have got the same linear momentum), then the scattering of the particles of each angular momentum could be considered independently of the particles of other angular momenta. Now, we have represented the incident beam of particles by a plane wave [Eq. (7.6)]. A plane wave is characterised by a definite linear momentum $\hbar\mathbf{k}$, but no definite angular momentum. In fact, a plane wave, being, in principle, of infinite extension, corresponds to impact parameters varying from zero to infinity. Correspondingly, the angular

9. The potential $V(\mathbf{r})$ corresponds to a single scattering centre ($n = 1$).
10. The impact parameter is the distance of the initial trajectory of the particle from the z-axis (which passes through the centre of the potential).

momenta contained in a plane wave also vary from zero to infinity. It is possible, therefore, to analyse a plane wave into an infinite number of components each of which corresponds to a definite angular momentum. Each of such components is called a *partial wave*, and the process of decomposing a plane wave into the partial waves is referred to as *partial wave analysis*.

Expansion of a Plane Wave in Terms of Partial Waves

We will now address ourselves to the problem of decomposing a plane wave into its partial waves. Now, the plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ is a solution of the free-particle Schrödinger Equation,

$$\frac{-\hbar^2}{2\mu}\Delta\psi = \frac{\hbar^2 k^2}{2\mu}\psi, \tag{7.10^1}$$

or,

$$(-\Delta - k^2)\psi(\mathbf{r}) = 0. \tag{7.11}$$

We will denote the normalized solutions of Eq. (7.11) by $v_{\mathbf{k}}(\mathbf{r})$. These are given by

$$v_{\mathbf{k}}(\mathbf{r}) = (2\pi)^{-3/2} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{7.12}$$

The orthonormal set $\{v_{\mathbf{k}}\}_\infty$, for k_x, k_y and k_z varying from $-\infty$ to $+\infty$, forms a basis for the infinite-dimensional Hilbert space of the particles. This basis defines the (linear) momentum representation for the system.

On the other hand, writing Eq (7.10^1) as

$$\hat{H}\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{7.10^2}$$

with

$$\hat{H} = -\frac{\hbar^2\Delta}{2\mu} = -\frac{\hbar^2}{2\mu}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) - \frac{\hat{\mathbf{L}}^2}{\hbar^2 r^2}\right], \tag{7.13}$$

we have,

$$[H, \hat{\mathbf{L}}^2] = \hat{0}, \tag{7.14a}$$

$$[H, \hat{L}_z] = \hat{0}, \tag{7.14b}$$

where, $\hat{\mathbf{L}}^2$ and $\hat{L}_z$ are the angular momentum operators given by Eqs. (5.43) and (5.42c). Eqs. (7.14a, b) imply that the eigenvectors of $\hat{H}$ are also the eigenvectors of $\hat{\mathbf{L}}^2$ and $\hat{L}_z$. But the eigenvectors of $\hat{\mathbf{L}}^2$ and $\hat{L}_z$ are the spherical harmonics $Y_{lm}(\theta,\phi)$ given by Eq. (5.55b):

$$\hat{\mathbf{L}}^2 Y_{lm}(\theta,\phi) = l(l+1)\hbar^2 Y_{lm}(\theta,\phi), \tag{7.15a}$$

$$\hat{L}_z Y_{lm}(\theta,\phi) = m\hbar Y_{lm}(\theta,\phi), \tag{7.15b}$$

where, the allowed values of l and m are given by Eqs. (5.49a, b). Thus,

$$\psi(\mathbf{r}) \equiv \psi_{klm}(r,\theta,\phi) = R_{k,l}(r)Y_{lm}(\theta,\phi), \tag{7.16^1}$$

where, $R_{k,l}(r)$ is a function of the radial co-ordinate only [(r, θ, ϕ) are the spherical co-ordinates of the particle, related to the cartesian co-ordinates through Eq. (4.107)]. Substituting from Eqs. (7.13), (7.16) and (7.15a) in Eq. (7.10^2), we get a differential equation for $R_{k,l}(r)$:

$$\left[\frac{d^2}{dr^2}+\frac{2}{r}\frac{d}{dr}+\left\{k^2-\frac{l(l+1)}{r^2}\right\}\right]R_{k,l}(r)=0, \tag{7.17}$$

where, $k^2=\frac{2\mu E}{\hbar^2}$. Eq. (7.17) is identical with Eq. (E.64). Therefore, the two independent solutions are the spherical Bessel function $j_l(kr)$ and the spherical Neumann function $n_l(kr)$. However, we see from Eq. (E.69b), that $n_l(kr)$ is not finite at $r = 0$. Since the wavefunction has to be finite everywhere, we have,

$$R_{k,l}(r) = j_l(kr). \tag{7.18}$$

Substituting this in Eq. (7.16^1) we get,

$$\psi_{klm}(r, \theta, \phi) = j_l(kr)Y_{lm}(\theta, \phi) \tag{7.16^2}$$

The infinite set $\{\psi_{klm}\}_\infty$ of orthogonal [See Eqs. (5.60) and (E.74)] vectors for $l = 0, 1, 2, \ldots +\infty$, $m = -l, -l+1, \ldots, +l$ and k varying from 0 to ∞, constitute a basis for the infinite-dimensional Hilbert space of the free particles. Each ψ_{klm} is called a *spherical* or *partial wave* as it is characterised by a definite angular momentum.

We have now two different bases, (7.12) and (7.16^2). The vectors in one should be expressible as a linear combination of the vectors in the other. That is,

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{l=0}^{\infty}\sum_{m=-l}^{+l} a_{lm}(\mathbf{r})j_l(kr)Y_{lm}(\theta, \phi) \tag{7.19^1}$$

The determination of the coefficients a_{lm} in (7.19^1) is facilitated by choosing the z-axis along $\mathbf{k}$. Then, $\mathbf{k}\cdot\mathbf{r} = kr\cos\theta = kz$, so that the L.H.S. of Eq. (7.19^1) becomes independent of the azimuthal angle ϕ. The R.H.S also should be, then, independent of ϕ, which requires that $m = 0$ [see Eq. (5.63)]. In view of the relation

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

where $P_l(x)$ is the Legendre polynomial of order l, Eq. (7.19^1) reduces to

$$e^{i\rho x} = \sum_{l=0}^{\infty} c_l j_l(\rho)P_l(x), \tag{7.19^2}$$

with

$$c_l = \sqrt{\frac{2l+1}{4\pi}}a_{l0};\ \rho = kr;\ x = \cos\theta. \tag{7.20}$$

Differentiating Eq. (7.19^2) with respect to ρ and then substituting for $\frac{dj_l}{d\rho}$ from Eq. (E.71b), we have,

$$ixe^{i\rho x} = \sum_{l=1}^{\infty} c_l \frac{l}{2l+1} j_{l-1} P_l(x) - \sum_{l=0}^{\infty} c_l \frac{l+1}{2l+1} j_{l+1} P_l(x). \tag{7.21a}$$

But,
$$ixe^{i\rho x} = i \sum_{l=0}^{\infty} c_l j_l x P_l(x)$$

$$= i \sum_{l=1}^{\infty} c_l j_l \frac{l}{2l+1} P_{l-1}(x) + i \sum_{l=0}^{\infty} c_l j_l \frac{l+1}{2l+1} P_{l+1}(x), \tag{7.21b}$$

where, expression (E.29a) for $xP_l(x)$ has been used. Equality of expressions (7.21a) and (7.21b) requires that coefficient of $P_l(x)$ in either expression be equal (since the P_l's for different values of l are linearly independent). That is,

$$\frac{l}{2l+1} c_l j_{l-1} - \frac{l+1}{2l+1} c_l j_{l+1} = i \left\{ \frac{l+1}{2l+3} c_{l+1} j_{l+1} + \frac{l}{2l-1} c_{l-1} j_{l-1} \right\} \tag{7.22}$$

Again, since the j_l's are linearly independent functions of ρ, coefficients of like orders of j_l on either side of (7.22) should be equal. From the coefficient of j_{l-1}, we have, the recurrence relation,

$$c_l = i \frac{2l+1}{2l-1} c_{l-1}, \; l \geq 1. \tag{7.23a}$$

This leads to the result:

$$c_l = i^l (2l+1) c_0. \tag{7.23b}$$

c_0 can be determined by putting $\rho = 0$ in Eq. (7.19^2) and using the values,

$$j_l(0) = \delta_{l0}, \text{ and } P_0(x) = 1.$$

We get,

$$c_0 = 1, \text{ so that,}$$

$$c_l = i^l (2l+1). \tag{7.23c}$$

Substituting in (7.19^2), we have the result,

$$\exp(ikr \cos\theta) \equiv \exp(ikz) = \sum_{l=0}^{\infty} (2l+1) i^l j_l(kr) P_l(\cos\theta). \tag{7.24}$$

This is known as *Bauer's Formula*. Using the spherical harmonic addition theorem [Eq. (5.152)], Eq. (7.24) can also be written as

$$\exp(i\mathbf{k}\cdot\mathbf{r}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} i^l j_l(kr) Y_{lm}^*(\theta_k, \phi_k) Y_{lm}(\theta_r, \phi_r), \tag{7.25}$$

where, (θ_k, ϕ_k) define the direction of $\mathbf{k}$ and (θ_r, ϕ_r) that of $\mathbf{r}$. Substituting the asymptotic value, (E.70a), for $j_l(kr)$ in Eq. (7.24), we have,

$$\exp(ikr\cos\theta) \underset{r\to\infty}{\sim} \frac{1}{kr} \sum_{l=0}^{\infty} (2l+1) i^l \sin\left(kr - \frac{l\pi}{2}\right) P_l(\cos\theta). \tag{7.26a}$$

Using the relation, $\sin x = \frac{1}{2i}(e^{ix} - e^{-ix})$, this can be written as

$$\exp(ikr\cos\theta) \underset{r\to\infty}{\sim} \frac{1}{2ikr}\sum_{l=0}^{\infty}(2l+1)[(-1)^{l+1}\exp(-ikr)+\exp(ikr)].P_l(\cos\theta)$$

(7.26b)

The term containing the factor $(e^{-ikr})/r$ represents an incoming spherical wave[11], while the term with the factor $(e^{ikr})/r$ represents an outgoing spherical wave (that is, particles that are approaching the scattering centre and those that are receding from the scattering centre, respectively).

Scattering by a Central Potential $V(r)$

In this case (of. Eq. (7.13)),

$$\hat{H} = -\frac{\hbar^2}{2\mu}\Delta + V(r), \tag{7.27}$$

so that Eqs. (7.14a, b) still hold good. The wavefunction is, therefore, of the same form as (7.16[1]). But now, the equation satisfied by $R_{k,l}(r)$ is,

$$\left[\frac{d^2}{dr^2}+\frac{2}{r}\frac{d}{dr}+\left\{k^2-U(r)-\frac{l(l+1)}{r^2}\right\}\right]R_{k,l}(r)=0, \tag{7.28}$$

where, $$U(r) = (2\mu/\hbar^2)V(r). \tag{7.29}$$

We assume that $V(r)\to 0$ as $r\to\infty$. Then, for large values of r, Eq. (7.28) reduces to the free-particle equation (7.17). Therefore, the solutions of (7.28) should asymptotically (that is, for a large values of r) approach the general solution of (7.17). Now, the general solution of Eq. (7.17) is a superposition of the spherical Bessel and the Neumann functions.[12] Thus,

$$\begin{aligned} R_{k,l}(r) &\underset{r\to\infty}{\sim} A_l j_l(kr) + B_l n_l(kr) \\ &\sim \frac{1}{kr}\left[A_l \sin\left(kr-\frac{1}{2}l\pi\right) - B_l\cos\left(kr-\frac{1}{2}/\pi\right)\right] \\ &\sim C_l \frac{\sin\left(kr-\frac{1}{2}l\pi+\delta_l\right)}{kr}, \end{aligned} \tag{7.30}$$

where, Eqs. (E.70a, b) have been used. Also, C_l and δ_l are related to A_l and B_l by

$$\tan\delta_l = -(B_l/A_l);\ A_l = C_l\cos\delta_l;\ -B_l = C_l\sin\delta_l. \tag{7.31}$$

The Schrödinger equation of the scattering problem can also be written as

---

11. For incoming waves, **k** and **r** are antiparallel so that $\mathbf{k}\cdot\mathbf{r} = kr\cos\pi = -kr$, whereas for outgoing ones, **k** and **r** are parallel.
12. The reason for excluding the spherical Neumann function in the solution of Eq. (7.17) is not valid here, since the solution (7.30) is restricted to regions outside the range of the potential and, thus, excludes the origin ($r = 0$).

$$\hat{H}\psi_k(\mathbf{r}) = \frac{\hbar^2 k^2}{2\mu}\psi_k(\mathbf{r}), \tag{7.32}$$

with $\hat{H}$ given by Eq. (7.27). The relations corresponding to (7.24) and (7.26a, b) are, then, given by

$$\psi_k(\mathbf{r}) = \sum_{l=0}^{\infty}(2l+1)i^l R_{k,l}(r)P_l(\cos\theta) \tag{7.33a}$$

$$\underset{r\to\infty}{\sim}\frac{1}{kr}\sum_{l=0}^{\infty}(2l+1)i^l C_l \sin\left(kr-\frac{1}{2}l\pi+\delta_l\right)P_l(\cos\theta)$$

$$\sim\frac{1}{2ikr}\sum_{l=0}^{\infty}(2l+1)\,C_l\exp(-i\delta_l)\,[(-1)^{l+1}\exp(-ikr)+$$

$$\exp(i2\delta_l)\exp(ikr)]P_l(\cos\theta). \tag{7.33b}$$

Now, the incoming part of the wavefunction in (7.33b) should be the same as in (7.26b), so that, we have,

$$C_l = \exp(i\delta_l), \tag{7.34}$$

and

$$\psi_k(\mathbf{r}) \underset{r\to\infty}{\sim} \frac{1}{2ikr}\sum_{l=0}^{\infty}(2l+1)\,[(-1)^{l+1}\exp(-ikr)$$

$$+\exp(2i\delta_l)\exp(ikr)]P_l(\cos\theta). \tag{7.33c}$$

For given l, the first term in (7.33c) represents an incoming spherical wave and the second term an outgoing spherical wave, both of the same intensity. The *phase* of the outgoing wave is, however, shifted relative to the phase of the corresponding wave in Eq. (7.26b) [that is, relative to the free-particle case] by the amount δ_l. δ_l is, therefore, called the *phase-shift*.[13]

We, thus, see that *the effect of the scattering potential is to shift the phase of each outgoing partial wave*.

The Scattering Amplitude

Now, the asymptotic wavefunction $\psi_k(\mathbf{r})$ for $r \to \infty$, should be of the form (7.8). Substituting for e^{ikz} in that equation from (7.26b), we have,

$$\psi_k(r) \underset{r\to\infty}{\sim} \frac{1}{2ikr}\sum_{l=0}^{\infty}(2l+1)\,[(-1)^{l+1}e^{-ikr}+e^{ikr}]P_l(\cos\theta)$$

$$+f_k(\Omega)\frac{e^{ikr}}{r}. \tag{7.35}$$

Comparing Eqs. (7.33c) and (7.35), we get,

---

13. That the phase shift is δ_l rather than $2\delta_l$, is seen from Eqs. (7.30), (7.18) and (E.70a).

$$f_k(\Omega) \equiv f_k(\theta) = \sum_{l=0}^{\infty} f_k^{(l)}(\theta), \tag{7.36}$$

where,
$$f_k^{(l)}(\theta) = \frac{2l+l}{k} \cdot \left(\frac{\exp(2i\delta_l) - 1}{2l} \right) P_l(\cos\theta).$$

$$= \{(2l+1)/k\} \exp(i\delta_l) \sin\delta_l P_l(\cos\theta). \tag{7.36a}$$

Thus, the scattering amplitude is independent of the azimuthal angle.

Using Eq. (7.9), we have,

$$\frac{d\sigma}{d\Omega} = |f_k(\Omega)|^2$$

$$= (1/k^2) \sum_{l=0}^{\infty} \sum_{l'=0}^{\infty} (2l+1)(2l'+1) \times$$

$$\exp[i(\delta_l - \delta_l')] \sin\delta_l \sin\delta_l' P_l(\cos\theta) P_l'(\cos\theta), \tag{7.37}$$

while, from Eqs. (7.3) and (7.37), we get, for the total cross-section, the expression,

$$\sigma = \int \frac{d\sigma}{d\Omega} d\Omega = \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \sin^2\delta_l$$

$$= \sum_{l=0}^{\infty} \sigma^{(l)}, \tag{7.38^1}$$

where, Eq. (E.31) has been used. $\sigma^{(l)}$ in Eq. (7.38^1) is the contribution to the total scattering cross-section from the lth partial wave. We note that

$$\sigma^{(l)} \leq (4\pi/k^2)(2l+1). \tag{7.39}$$

From Eqs. (7.36), (7.36a) and (7.38^1), we have,

$$\sigma = (4\pi/k)\ \mathrm{Im}\{f_k(\theta = 0)\}, \tag{7.38^2}$$

where, Im $\{f_k(\theta = 0)\}$ is the imaginary part of the *forward* scattering amplitude. Now, the total cross-section represents the loss of intensity suffered by the incident beam (in the direction $\theta = 0$) resulting from the fact that some particles have been deflected away from the incident direction. Eq. (7.38^2), then, states that this loss of intensity is represented by the imaginary part of the scattering amplitude in the forward direction. In analogy with a similar case in optics, where the imaginary part of the complex index of refraction is related to the absorption cross-section for light in the medium, Eq. (7.38^2) is called the *Optical Theorem* (also, sometimes, referred to as the *Bohr-Peierls-Placzek formula*).

Dependence of δ_l on V

The phase shift will, obviously, depend on the potential (It is zero when $V(r) = 0$). This dependence can be derived from the radial equations (7.17) and (7.28):

Put $R_{k,l}(r) = \{v_{k,l}(r)\}/r$, in Eq. (7.17), and $R_{k,l}(r) = \{u_{k,l}(r)\}/r$ in Eq. (7.28). Then, the radial equations reduce to:

$$\left[\frac{d^2}{dr^2}+k^2-\frac{l(l+1)}{r^2}\right]v_{k,l}(r)=0, \tag{7.40}$$

and

$$\left[\frac{d^2}{dr^2}+k^2-\left\{U(r)+\frac{l(l+1)}{r^2}\right\}\right]u_{k,l}(r)=0. \tag{7.41}$$

Multiply Eq. (7.41) from the left by $v_{k,l}(r)$ and then integrate over r from 0 to ∞. Similarly, multiply Eq. (7.40) by $u_{k,l}(r)$ and integrate from 0 to ∞. Subtract the latter result from the former. We get,

$$\int_0^\infty v_{k,l}\frac{d^2u_{k,l}}{dr^2}dr-\int_0^\infty u_{k,l}\frac{d^2v_{k,l}}{dr^2}dr=\int_0^\infty v_{k,l}U(r)u_{k,l}dr.$$

Integrating the L.H.S. by parts, we have,

$$\left[v_{k,l}\frac{du_{k,l}}{dr}-u_{k,l}\frac{dv_{k,l}}{dr}\right]_0^\infty=\int_0^\infty v_{k,l}U(x)u_{k,l}dr. \tag{7.42}$$

Now,

$$v_{k,l}(r)=rj_l(kr)\underset{r\to\infty}{\sim}(1/k)\sin\left(kr-\frac{1}{2}l\pi\right), \tag{7.43a}$$

$$u_{k,l}(r)=rR_{k,l}(r)\underset{r\to\infty}{\sim}(C_l/k)\sin\left(kr-\frac{1}{2}l\pi+\delta_l\right), \tag{7.43b}$$

where, Eqs. (7.18) and (7.30) have been used, and C_l is given by Eq. (7.34). Also, since, $R_{k,l}(r)$ should be finite at $r=0$,

$$v_{k,l}(0)=u_{k,l}(0)=0. \tag{7.44}$$

Substituting from Eqs. (7.43 a, b) and (7.44) in Eq. (7.42), we get,

$$-(C_l/k)\sin\delta_l=\int_0^\infty rj_l(kr)U(r)u_{k,l}(r)dr,$$

or

$$\sin\delta_l=-k\int_0^\infty rj_l(kr)U(r)\{u_{k,l}(r)/C_l\}dr. \tag{7.45^1}$$

Suppose now, that $U(r)$ is infinitesimal. In this case, $\{u_{k,l}(r)/C_l\}$ will differ little from the corresponding free-particle value. That is, $\{u_{k,l}(r)/C_l\}\approx rj_l(kr)$, and (since δ_l would be small),

$$\sin\delta_l\approx\delta_l=-k\int_0^\infty U(r)\{j_l(kr)\}^2r^2dr. \tag{7.45^2}$$

Since $r^2j_l^2(kr)$ is always positive, the R.H.S. has the sign opposite to that of $U(r)$. Thus, δ_l is positive for a negative potential (attractive force) and δ_l is negative for

a positive potential (repulsive force). Since a finite potential can be built up by successively adding infinitesimal potentials, the above statement regarding the relative signs of δ_l and $V(r)$ holds good for the case of finite potentials also. This result could have been anticipated, for, the particles would be accelerated and would, thus, gain phase in an attractive field whereas they would be slowed down (lose phase) in a repulsive field.

Dependence of δ_l on the Angular Momentum and the Energy

Eq. (7.41) can be written as

$$\left[\frac{d^2}{dr^2}+k^2-\frac{2\mu}{\hbar^2}V_{\text{eff}}(r)\right]u_{k,l}(r)=0, \tag{7.46}$$

where

$$V_{\text{eff}}(r)=V(r)+\frac{l(l+1)\hbar^2}{2\mu r^2}, \tag{7.47}$$

could be regarded as the 'effective potential'. Eq. (7.46), thus, represents a particle of energy $E=\dfrac{\hbar^2k^2}{2\mu}$, moving under the influence of a potential[14] $V_{\text{eff}}(r)$. Fig. 7.3 shows the variation of $V_{\text{eff}}(r)$ and $\dfrac{l(l+1)\hbar^2}{2\mu r^2}$ for a typical potential $V(r)$. r_1 represents the classical *turning point* (the closest approach), and is given by

$$E=\frac{\hbar^2k^2}{2\mu}=V_{\text{eff}}(r_1)=V(r_1)+\frac{l(l+1)\hbar^2}{2\mu r_1^2}. \tag{7.48}$$

For a given value of E, r_1 can be made as large as desired by choosing a sufficiently large value of l. For $r<r_1, k^2-(2\mu/\hbar^2)V_{\text{eff}}(r)$ in Eq. (7.46) would be negative, so that the wavefunction $u_{k,l}(r)$ would be exponential. Since further , $u_{k,l}(0) = 0$, $u_{k,l}(r)$ in the region $r<r_1$ is as shown in Fig. 7.3. Therefore, the product $u_{k,l}(r)V(r)$ will be small for all values of $r<r_1$, provided the range r_0 of the potential is small compared with r_1. Under this condition, $V(r)$ would be negligible for $r>r_1$, so that, since $u_{k,l}(r)$ is oscillatory in this region, the product

---

14. The term $\dfrac{l(l+1)\hbar^2}{2\mu r^2}$ in Eq. (7.47) represents the centrifugal force experienced by a particle in its orbital motion. For, if F_c represents the centrifugal force, we have, $F_c=-\dfrac{dV_c}{dr}=-\mu\omega^2 r$, where, ω is the angular velocity of the particle. Hence, $V_c(r)=\frac{1}{2}\mu\omega^2r^2=\frac{1}{2}\mathcal{J}\omega^2=(\mathrm{L}^2/2\mathcal{J})$, where $\mathcal{J}=\mu r^2$, is the moment of inertia of the particle, and $\mathrm{L}=\mathcal{J}\omega$, the angular momentum.

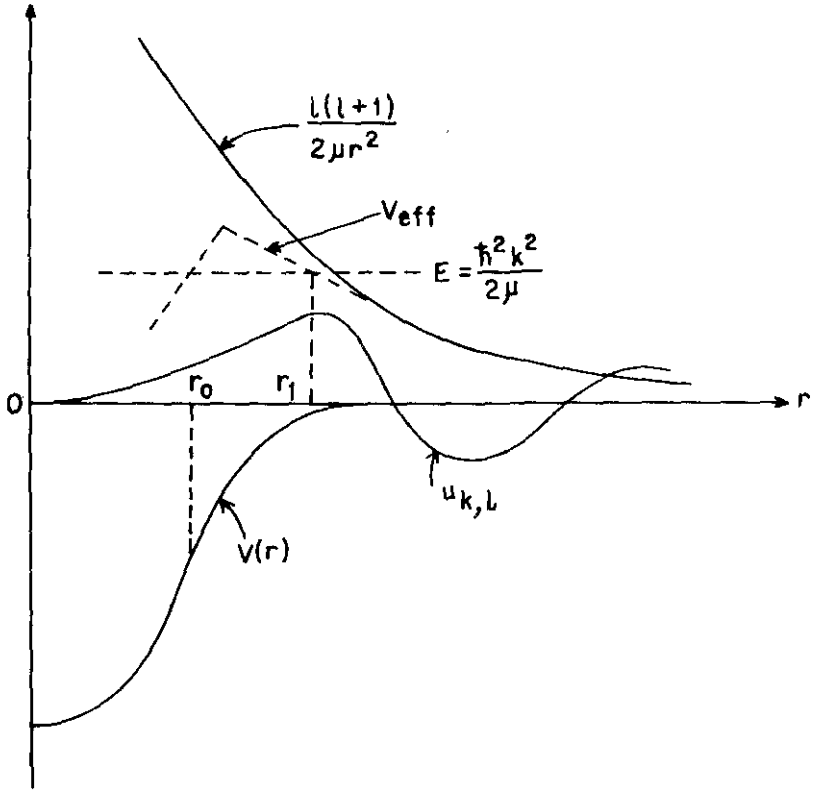


Fig. 7.3 Diagram showing the conditions for the product $V(r)u_{k,l}(r)$ in Eq. (7.45[1]) to be small throughout the range $r = 0$ to $r = \infty$.

$u_{k,l}(r)V(r)$ would be small for all values of $r > r_1$. Thus, we see that the product $u_{k,l}(r)V(r)$ would be negligible everywhere provided $r_1 \gg r_0$. And under this condition, we have, from Eq. (7.45[1]), the result $\delta_l \approx 0$.

Now, the condition $r_1 \gg r_0$, is equivalent to the condition,

$$\frac{l(l+1)\hbar^2}{2\mu r_1^2} \gg V(r_1),$$

or

$$V_{\text{eff}}(r_1) \approx \frac{l(l+1)\hbar^2}{2\mu r_1^2}.$$

i.e.

$$E = \frac{\hbar^2 k^2}{2\mu} \approx \frac{l(l+1)\hbar^2}{2\mu r_1^2},$$

or

$$l(l+1) \approx l^2 \approx k^2 r_1^2 \gg k^2 r_0^2$$

That is,

$$l \gg kr_0. \tag{7.49}$$

Thus, whenever condition (7.49) is satisfied, we have, $\delta_l \approx 0$. Classically, this result corresponds to the fact that particles with impact parameters larger than the range of the potential are not scattered.

Now, condition (7.49) would be satisfied by any non-zero value of l if k (and hence the energy) is small enough. Thus,

$$Lt_{E\to 0}\delta_l = 0, \; l \neq 0. \tag{7.50}$$

Also, when the energy is small, the phase shifts would be small and would, therefore, be given by Eq. (7.45[2]). Substituting expression (E.69a) for $j_l(kr)$ in (7.45[2]), we get,

$$\sin \delta_l \approx -\alpha_l^2 \left(\frac{2\mu}{\hbar^2}\right)^{l+3/2} E^{l+1/2} \int_0^\infty V(r) r^{2l+2} dr, \tag{7.51}$$

$$\approx \delta_l, \text{ for } l \neq 0,$$

where $$\alpha_l = \frac{1}{(2l+1)!!}.$$

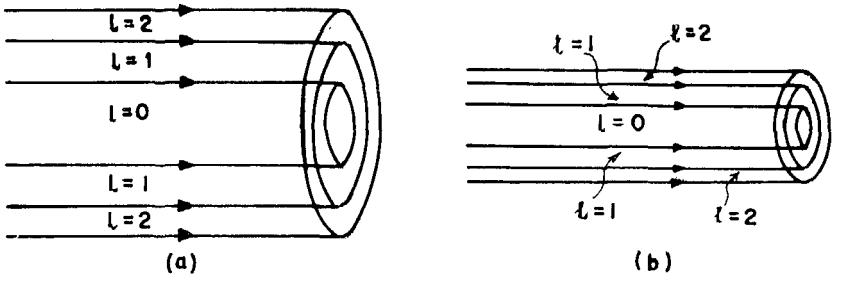


Fig. 7.4 The space occupied by different partial waves, (a) For E~0, (b) for larger E.

Thus, the phase shift varies as $E^{l+1/2}$ for small values of E. When the energy is practically zero, only the s-wave ($l = 0$) phase shift is non-zero. As the energy is increased, higher and higher partial waves come under the influence of the scattering potential. This phenomenon is illustrated in Fig. 7.4 where the space occupied by a given partial wave is shown to shrink as the energy is increased.

Zero-Energy Scattering: The Scattering Length

The method of partial waves is useful in practice only if the series (7.36) and (7.38[1]) converge rapidly so that all but a few of the terms in the series can be neglected. From the discussion leading to Eqs. (7.49) and (7.50), it follows that the necessary condition is that the energy be small and the potential be of short-range. That is, the method of partial waves is suited only for low energy scattering.

If the energy is so low that only particles with $l = 0$ are scattered, then the theory is especially simple. In this case, only δ_0 (the 's-wave phase shift') would different from zero. Then, from Eq. (7.36) and (7.36a), we have,

$$f_k(\theta) = f_k^{(0)}(\theta) = \frac{1}{k} e^{i\delta_0} \sin \delta_0, \tag{7.52}$$

which is independent of the scattering angle. The limiting value of the energy for which Eq. (7.52) holds good, is called 'zero-energy'. Thus, the angular distribution of the scattered particles at zero energy, is independent of the scattering angle. In other words, scattering is *isotropic*[15].

The negative of the scattering amplitude in the zero-energy limit is called the *scattering length*[16] and is denoted by 'a'. Thus,

$$a = Lt_{E \to 0}[-f_k(\theta)] = -\frac{1}{k} e^{i\delta_0} \sin \delta_0. \tag{7.53}$$

In terms of 'a' the zero-energy total scattering cross-section is given by (see Eq. (7.38[1])),

$$\sigma = 4\pi |f_k(\theta)|^2 = 4\pi a^2. \tag{7.54}$$

Geometrical Interpretation of Scattering Length

The radial equation (7.41) for $l = 0$, reduces to

$$\left[\frac{d^2}{dr^2} + k^2 - U(r)\right] u_{k,0}(r) = 0. \tag{7.55}$$

In the zero-energy limit and for $r \gg r_0$, this becomes,

$$\frac{d^2 u_{k,0}}{dr^2} \equiv \frac{d^2 u_0}{dr^2} \approx 0. \tag{7.55a}$$

Thus, the asymptotic value of $u_0(r)$ is given by,

$$v_0(r) \equiv \{u_0(r)\}_{r \gg r_0} = br + c, \tag{7.56}$$

where, b and c are constants.

But,

$$u_{k,0} = rR_{k,0}(r) \underset{r \gg r_0}{\sim} re^{i\mathbf{k}\cdot\mathbf{r}} + f_k(\theta) e^{ikr}$$

$$\underset{k \to 0}{\approx} (r - a) = v_0(r), \tag{7.57a}$$

where use has been made of expressions (7.33a), (7.8) and (7.53). In Eq. (7.57a), a normalization constant is arbitrary, so that, we may write,

$$v_0(r) = \alpha(r - a) \tag{7.57b}$$

Further, choosing the normalization, $|\alpha| = 1$, we have,

$$v_0(r) = \pm(r - a), \; r \gg r_0. \tag{7.57c}$$

---

15. Isotropy is characteristic of the s-state. Since the angular momentum is zero, there is nothing to fix a direction in space that is common to both the incident and the scattered particles.
16. This definition ensures that the scattering length is positive for bound states. It also enables us to interpret 'a' as the radius of a hard ('impenetrable') sphere which would cause the same amount of scattering as the bound system.

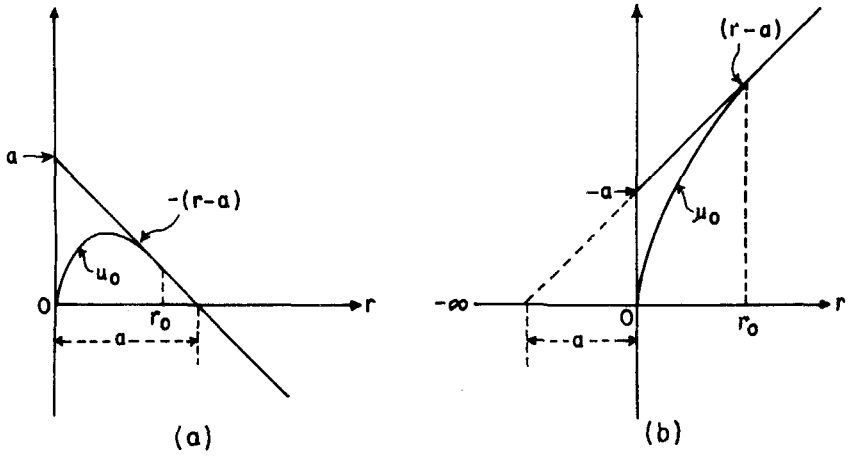


Fig. 7.5 Geometrical determination of the scattering length for a potential of range r_0.

Now, Eq. (7.56) is the equation of a straight line with slope 'b' and intercept on the u_0-axis equal to 'c'. Comparing Eqs. (7.56) and (7.57c), we see that 'a' is the intercept of the straight line $u_0(r)$ on the u_0-axis when the slope of the straight line is negative as in Fig. (7.5a), and is the negative of this intercept when the slope is positive as in Fig. (7.5b). In either case, 'a' is equal to the intercept of the straight line on the r-axis[17] as is seen by putting $u_0(r) = 0$ in Eq. (7.57c).

When the potential has a sharp boundary so that $V(r) = 0$, for $r > r_0$, the application of the above geometrical method for obtaining 'a' is especially simple. The continuity of the wavefunction requires that the interior wavefunction $u_0(r)$ match the exterior wavefunction $v_0(r)$ at $r = r_0$. That is, $u_0(r_0) = v_0(r_0)$, and $u_0'(r_0) = v_0'(r_0)$. Thus, $v_0(r)$ would be the tangent to $u_0(r)$ at $r = r_0$ (as shown in Fig. 7.5). A few interesting cases are shown in Fig. 7.6. Fig. 7.6(c) corresponds to the case of an impenetrable sphere of radius r_0. Since the wavefunction should be zero inside such a sphere, the continuity of the wavefunction at $r = r_0$ requires that it be zero at the surface of the sphere also. The result for this case suggests (see Footnote 16) the possibility of interpreting the scattering length as the radius of a hard sphere which is equivalent to the actual physical system as far as scattering at low energies is concerned[18]. From Eq. (7.54), we have the interesting result, $\sigma = 4\pi r_0^2$. That is, the total scattering cross-section is four times

17. It should be noted that the straight line $(r - a)$ has really no intercept on the r-axis when $u_0(r)$ is as shown in Fig. 7.5(b), since r has no negative values. However, purely as a geometrical procedure, we can extend the r-axis to the negative side. Then, Eq. (7.57c) shows that 'a' is equal, both in magnitude and sign, to the intercept of the asymptotic straight line on this extended r-axis.
18. Since the cross-section is proportional to the square of the radius, a negative value for the radius cannot be ruled out. As seen in Fig. 7.6(d), a negative value corresponds to the case in which the potential is neither repulsive enough to have a hard core nor attractive enough to form a bound system (with a real radius).

the geometrical cross-section! Actually, it is not this result that is unphysical (as, at first, we would be tempted to exclaim), but the assumption leading to it. That assumption consists in identifying the surface of the sphere with an *infinite* potential barrier. Infinite potential barriers are unphysical as there are no systems that are perfectly rigid. What the result implies is that, if the wavefunction is zero within a sphere of radius r_0, then the radius of the actual (physical) sphere is at least $2r_0$ (see the discussion on barrier penetration under WKB Approximation, Section 8.1).

Whereas a negative scattering length [Fig. 7.6(d)] implies that the potential is not capable of binding, a positive scattering length need not signify a bound state Fig. 7.6[(b) and (c)]. However, for bound states, the scattering length is necessarily positive as the wavefunction should fall off exponentially outside the range of the potential. Thus, in the case of scattering of neutrons by protons, the scattering length is -23.7×10^{-13} cm when the spins of the two particles are antiparallel (singlet state) whereas it is $+5.4 \times 10^{-13}$ cm when their spins are parallel (triplet state). It is established by other independent experiments that the bound state of the neutron-proton system (the deuteron) is indeed a triplet state. The large negative value of the scattering length in the singlet state, however, shows that the neutron-proton interaction in the singlet state misses binding only marginally ($\delta_0 \leq \pi/2$).

Also, since a new bound state appears whenever δ_0 crosses an odd multiple of $\pi/2$, the number n of bound states is related to the phase shift δ_0 by the inequality, $\left(n - \frac{1}{2}\right)\pi < \delta_0 < \left(n + \frac{1}{2}\right)\pi$.

(f) represents the *Ramsauer-Townsend Effect*, the very low minimum observed in the cross-section for the scattering of electrons by rare-gas atoms at about 0.7 eV bombarding energy.

Problem 7.1: Derive the relationship, $\delta_0 = -ka + n\pi$, $n = 0, 1, 2, \ldots$.

Scattering by a Square-Well Potential: Effective Range

The preceding discussion shows that zero-energy scattering can be characterised by just one parameter—the scattering length. The differential scattering cross-section is independent of both the energy and the scattering angle. The latter feature is a consequence of the fact that only s-waves (particles with angular momentum $l = 0$, which are insensitive to the *shape* of the potential) are scattered. The energy-independence follows from Eq. (7.51) which shows that

$$\left| \frac{\sin \delta_0}{k} \right|^2_{k \to 0} = |f_k(\theta)|^2_{k \to 0} = \left(\frac{d\sigma}{d\Omega} \right)_{k \to 0}$$

[see Eqs. (7.52) and (7.9)] is independent of the energy.

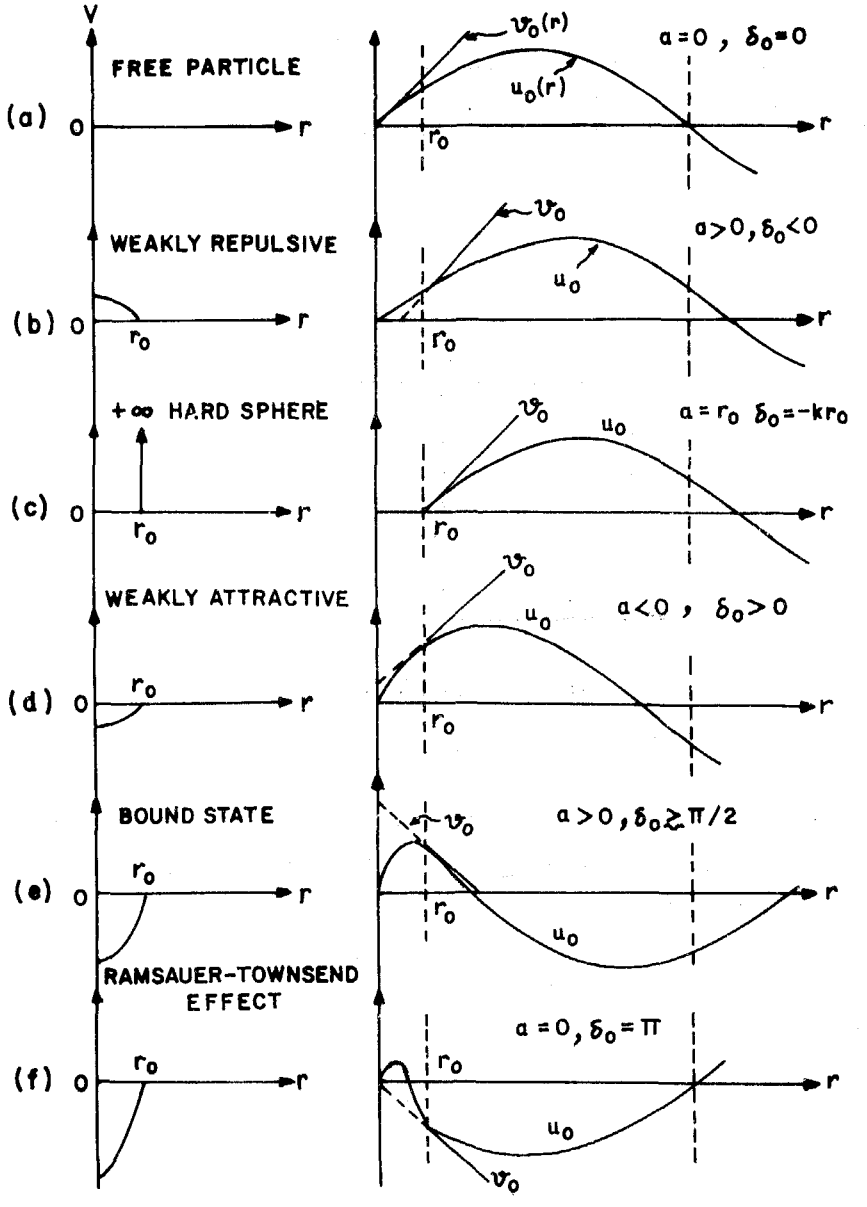


Fig. 7.6 Variation of the scattering length and the phase shift with the potential. $v_0 = \pm(r - a)$, $u_0 \sim \sin(kr + \delta_0)$.

As the energy is increased, more and more of the partial waves will begin to get scattered making the scattering dependent on both the energy and the scattering angle. If the energy is only slightly higher than what is termed zero-energy, the scattering would still be confined to the *s*-waves (and, hence, angle-independent), but the energy-dependence would make its appearance [Formula

(7.51), which is based on the approximation (E.69a), would no longer be valid]. This energy-dependence of the scattering cross-section at low energies can be described in terms of a parameter called *effective range*. We will illustrate this in the case of scattering by a short-range, attractive, square-well potential.

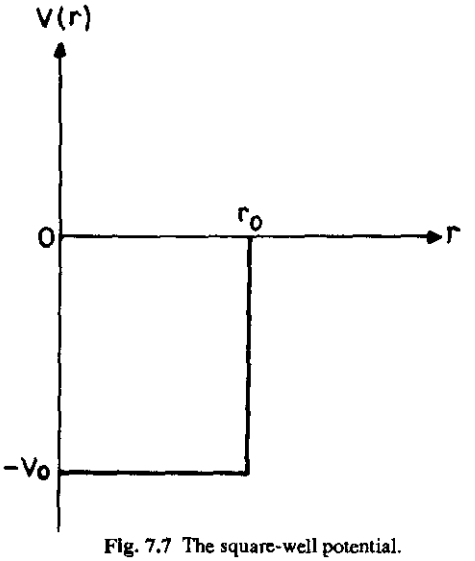


Fig. 7.7 The square-well potential.

Let (see Fig. 7.7)

$$V(r) = -V_0, \text{ for } r < r_0$$

$$= 0, \text{ for } r > r_0. \tag{7.58}$$

Also, let $u(r)$ and $v(r)$ represent $u_{k,0}(r)$ of Eq. (7.55) for $r < r_0$ and $r > r_0$, respectively. Then, we have,

$$\left(\frac{d^2}{dr^2} + \kappa^2\right) u(r) = 0, \quad r < r_0 \tag{7.59a}$$

and

$$\left(\frac{d^2}{dr^2} + k^2\right) v(r) = 0, \quad r > r_0, \tag{7.59b}$$

where

$$\kappa^2 = k^2 + (2\mu/\hbar^2) V_0. \tag{7.60}$$

The boundary conditions to be satisfied by $u(r)$ and $v(r)$ are:

$$u(0) = 0; \; u(r_0) = v(r_0); \; \frac{du}{dr}\Big|_{r=r_0} = \frac{dv}{dr}\Big|_{r=r_0}. \tag{7.61}$$

Hence the solutions are:

$$u(r) = A \sin \kappa r, \quad r < r_0, \tag{7.62}$$

and

$$v(r) = B \sin(kr + \delta), \quad r > r_0. \tag{7.63^1}$$

But [see Eqs. (7.43b) and (7.34)],

$$v(r) = \frac{e^{i\delta}}{k} \sin(kr + \delta),$$

where $\delta \equiv \delta_0$. In place of this normalization, we adopt the normalization[19],

$$v(0) = 1. \tag{7.63^2}$$

Then

$$B = \frac{1}{\sin\delta}, \tag{7.64a}$$

and, from the second of the boundary conditions (7.61),

$$A = \frac{\sin(kr_0 + \delta)}{\sin\delta \sin\kappa r_0}. \tag{7.64b}$$

The normalization (7.63^2) makes both $v(r)$ and $u(r)$ dimensionless, so that, in place of (7.57c), we have now the relation,

$$v(r) \underset{k\to 0}{\sim} \frac{\sin(kr + \delta)}{\sin\delta}\Big|_{k\to 0} = \left\{\frac{\sin k(r-a)}{\sin(-ka)}\right\}_{k\to 0}$$

$$= 1 - \frac{r}{a}, \tag{7.65}$$

where the relation $\delta = -ka + n\pi$ (Problem 7.1) is used.

The last two of the boundary conditions require that the logarithmic derivative be continuous at $r = r_0$. That is,

$$\frac{1}{u}\frac{du}{dr}\Big|_{r=r_0} = \frac{1}{v}\frac{dv}{dr}\Big|_{r=r_0}. \tag{7.61a}$$

Then, from (7.62) and (7.63^1), we get,

$$\kappa \cot \kappa r_0 = k \cot (kr_0 + \delta). \tag{7.66}$$

Problem 7.2: Using Eq. (7.66), show that

$$a = r_0\left(1 - \frac{\tan Kr_0}{Kr_0}\right),$$

where $K^2 = (2\mu/\hbar^2)V_0$.

Our objective is to find the dependence of the cross-section on the energy. Therefore, let us consider Eqs. (7.59a, b) for two different values of k, say k_1 and k_2. We have, the four equations.

---

19. Since $v(r)$ corresponds to the actual wavefunction only outside the range of the potential, the value $v(0) = 1$, does not affect the value of the actual wavefunction at $r = 0$.

$$\left[\frac{d^2}{dr^2}+k_1^2+(2\mu/\hbar^2)V_0\right]u_1(r)=0, \tag{7.67a}$$

$$\left[\frac{d^2}{dr^2}+k_2^2+(2\mu/\hbar^2)V_0\right]u_2(r)=0. \tag{7.67b}$$

$$\left[\frac{d^2}{dr^2}+k_1^2\right]v_1(r)=0, \tag{7.68a}$$

$$\left(\frac{d^2}{dr^2}+k_2^2\right)v_2(r)=0. \tag{7.68b}$$

Multiplying Eq. (7.67a) from the left by u_2 and Eq. (7.67b) by u_1, subtracting the latter result from the former and integrating over r, we get,

$$\left(u_2\frac{du_1}{dr}-u_1\frac{du_2}{dr}\right)\Big|_0^\infty=(k_2^2-k_1^2)\int_0^\infty u_1u_2dr,$$

or,

$$W(u_2,u_1)|_0^\infty=(k_2^2-k_1^2)\int_0^\infty u_2u_1dr, \tag{7.69}$$

where $W(u_2,u_1)$ is the *Wronskian* of u_2 and u_1.

Similarly, from Eqs. (7.68a, b), we have,

$$W(v_2,v_1)|_0^\infty=(k_2^2-k_1^2)\int_0^\infty v_2v_1dr. \tag{7.70}$$

From Eqs. (7 69) and (7.70), we get,

$$[W(u_2,u_1)-W(v_2,v_1)]_0^\infty=(k_2^2-k_1^2)\int_0^\infty(u_1u_2-v_1v_2)dr \tag{7.71}$$

Now,

$$W(u_2,u_1)|_0=0, \text{ since } u_1(0)=u_2(0)=0.$$

Also,

$W(u_2,u_1)|_\infty=W(v_2,v_1)|_\infty$ since, for large values of r, the interior and the exterior solutions should coincide with each other. Thus, Eq. (7.71) reduces to

$$W(v_1,v_2)|_0=(k_2^2-k_1^2)\int_0^\infty(v_1v_2-u_1u_2)dr. \tag{7.72^1}$$

But, from Eqs. (7.63^{1,2}) and (7.64a) , we have,

$$v_1(r)\frac{dv_2}{dr}\Big|_{r=0}=k_2\cot\delta_2, \tag{7.73}$$

so that,

$$W(v_1,v_2)|_0=k_2\cot\delta_2-k_1\cot\delta_1. \tag{7.74}$$

Substituting (7.74) in (7.72^1), we get,

$$k_2\cot\delta_2-k_1\cot\delta_1=(k_2^2-k_1^2)\int(v_1v_2-u_1u_2)dr. \tag{7.72^2}$$

Consider two energies which differ only by an infinitesimal amount. Then,

$$k_2^2 - k_1^2 = \Delta(k^2),$$

and

$$k_2 \cot \delta_2 - k_1 \cot \delta_1 = \Delta(k \cot \delta).$$

Substituting these in Eq. (7.72^2) and then taking the limit $\Delta(k^2) \to 0$, we get,

$$\underset{\Delta k^2 \to 0}{\text{Lt}} \frac{\Delta(k \cot \delta)}{\Delta(k^2)} \equiv \frac{d}{dk^2}(k \cot \delta) = \int_0^\infty (v^2 - u^2) dr. \tag{7.75}$$

Integrating over k^2 between limits 0 to k^2, we get,

$$k \cot \delta = (k \cot \delta)_{k=0} + \int_0^{k^2} dk^2 \int_0^\infty (v^2 - u^2) dr. \tag{7.76^1}$$

Now, $(k \cot \delta)_{k=0} = \dfrac{d}{dr} v(r)\Big|_{k=0}$

$$= \frac{d}{dr}\left(1 - \frac{r}{a}\right) = -\frac{1}{a}, \text{ (from Eq. (7.65))} \tag{7.77}$$

Also, $(v^2 - u^2)$ is dimensionless, so that $\int_0^\infty (v^2 - u^2) dr$ has the dimension of length. Moreover, this length is approximately independent of k for small values of k^2, and is of the order of the range of the potential [since, for $r < r_0$, the behaviour of the wavefunction is mainly determined by the potential when $k^2 \ll (2\mu/\hbar^2)V_0$, whereas for $r > r_0$, $(v^2 - u^2)$ is practically zero]. Thus,

$$k \cot \delta \approx -\frac{1}{a} + \left\{\int_0^\infty (v^2 - u^2)_{k=0} dr\right\} k^2$$

$$\approx -\frac{1}{a} + \frac{1}{2} r_{\text{eff}} k^2, \tag{7.76^2}$$

where

$$r_{\text{eff}} = 2\int_0^\infty (v^2 - u^2)_{k=0} dr, \tag{7.78}$$

is called the effective range[20] of the potential. We have, from Eq. (7.76^2),

$$\sin^2 \delta = \frac{1}{1 + \cot^2 \delta} = \frac{k^2}{k^2 + \left(\frac{1}{2} r_{\text{eff}} k^2 - \frac{1}{a}\right)^2},$$

and

$$\sigma = \frac{4\pi}{k^2} \sin^2 \delta = \frac{4\pi}{k^2 + \left(\frac{1}{2} r_{\text{eff}} k^2 - \frac{1}{a}\right)^2}. \tag{7.79}$$

---

20. With our normalization (7.64a, b), $(v^2 - u^2)_{k=0}$ is always positive, so that r_{eff} is a positive quantity.

This formula gives the energy-dependence of the total (s-wave) cross-section at low energies. It shows that low-energy scattering by a short range attractive potential can be described in terms of two parameters – the scattering length and the effective range. The absolute value of 'a' can be determined from experiments at zero energy [Formula (7.54), which also agrees with (7.79) when $k = 0$]. The sign of 'a' and the value of r_{eff} can be determined from the measured cross-sections at higher energies and formula (7.79).

Problem 7.3: Determine the values of $V_0 r_0^0$ for a square-well potential such that the cross-section at zero energy is zero.

Problem 7.4: If γ_l represents the logarithmic derivative of the radial wavefunction at the boundary $r = r_0$:

$$\gamma_l = \left\{ \frac{1}{R_l} \left(\frac{dR_l}{dr} \right) / R_l \right\}_{r=r_0};$$

show that the phase shift is given by

$$\tan \delta_l = \frac{k j_l'(kr_0) - \gamma_l j_l(kr_0)}{k n_l'(kr_0) - \gamma_l n_l(kr_0)}.$$

Deduce from this the relationship $\delta_0 = -kr_0$ for low-energy scattering by a hard sphere of radius r_0.

Resonance Scattering

The conclusion that only s-waves are scattered near zero energy, is based on the assumption that the phase shift and the cross-section vary slowly and smoothly with energy. However, large variations in the cross sections over a small interval of energy do some times occur (depending on the potential). Such large variations in the cross-section are attributed to the fact that a certain partial wave is *in resonance* with the potential near zero energy. The phenomenon is, therefore, called *resonance scattering*. Naturally, it is the partial wave which is in resonance rather than the s-wave, that dominates scattering near the resonance energy.

An understanding of the phenomenon of resonance could be gained from the following considerations: In terms of the logarithmic derivative

$$\gamma_l(k) = [d R_{k,l}/dr / R_{k,l}(r)]_{r=r_0} = \left[du_{k,l}/dr / u_{k,l}(r) \right]_{r=r_0} - \frac{1}{r_0},$$

the phase shift δ_l is given by (see Problem 7.4),

$$\tan \delta_l = \frac{k j_l'(\rho_0) - \gamma_l j_l(\rho_0)}{k n_l'(\rho_0) - \gamma_l n_l(\rho_0)}, \tag{7.80}$$

where $\rho_0 = kr_0$; and r_0 = the range of the potential.

Writing, $i \tan \delta_l = \dfrac{i \sin \delta_l}{\cos \delta_l}$, and using the relationship, $\dfrac{a+b}{a-b} = \dfrac{c+d}{c-d}$, when $\dfrac{a}{b} = \dfrac{c}{d}$,

we obtain from (7.80) the equation,

$$\exp(2i\delta_l) = \exp\{2i(\xi_l + \zeta_l)\}, \tag{7.81}$$

where

$$\exp(2i\xi_l) = -\frac{j_l(\rho_0) - i n_l(\rho_0)}{j_l(\rho_0) + i n_l(\rho_0)}, \tag{7.82a}$$

$$\exp(2i\zeta_l) = \frac{\beta_l - \Delta_l + i s_l}{\beta_l - \Delta_l - i s_l}, \tag{7.82b}$$

with

$$\beta_l(k) = r_0 \gamma_l(k), \tag{7.83a}$$

$$\Delta_l(\rho_0) = \frac{\rho_0(j_l j'_l + n_l n'_l)}{j_l^2 + n_l^2}, \tag{7.83b}$$

and

$$s_l(\rho_0) = \frac{1}{\rho_0(j_l^2 + n_l^2)}. \tag{7.83c}$$

Here, the argument of the spherical Bessel and Neumann functions is ρ_0. Also, Eq. (E.73) has been used in obtaining (7.83c).

Eq. (7.82a) which is equivalent to the relation,

$$\tan \xi_l = \frac{j_l(kr_0)}{n_l(kr_0)}, \tag{7.84^1}$$

shows that ξ_l is real. Similarly, Δ_l, s_l and, hence,

$$\zeta_l = \tan^{-1}\{s_l/(\beta_l - \Delta_l)\}, \tag{7.85^1}$$

are also real.

Eq. (7.81) shows that the phase shift δ_l can be written as the sum of two terms:

$$\delta_l = \xi_l + \zeta_l. \tag{7.86}$$

Of the two, ξ_l is a slowly varying function of energy. In fact, substituting from Eqs. (E.69a, b) in (7.84^1), we have,

$$\tan \xi_l \underset{kr_0 \to 0}{\sim} -\frac{(kr_0)^{2l+1}}{(2l+1)\{(2l-1)!!\}^2}. \tag{7.84^2}$$

Thus, at low energies, ξ_l goes to zero as $E^{l+1/2}$ [c.f. Eq. (7.51)]. Also, we see from Eq. (7.82b) that $\zeta_l \to 0$ as $\beta_l \to \infty$. Therefore, ξ_l is the total phase shift whenever β_l is infinite. Now, an infinite β_l corresponds to a hard sphere [see Fig. 7.6(c)]. For this reason, ξ_l is called the *hard-sphere phase shift*.

As is seen from Eq. (7.85[1]) the other term, ζ_l, depends very sensitively on the energy because of the factor $(\beta_l - \Delta_l)$ in the denominator[21]. Obviously, a sudden increase in the value of $\tan \zeta_l$ occurs for $\beta_l \approx \Delta_l$. If the energy is small enough, ξ_l would be negligible compared with ζ_l at this energy, so that $\delta_l \approx \zeta_l$. Also, the phase shifts corresponding to the partial waves other than the one for which $\beta_l \approx \Delta_l$ can be neglected. Then,

$$\sigma \approx \sigma^{(l)} = \frac{4\pi}{k^2}(2l+1)\sin^2\delta_l$$

$$\approx \frac{4\pi(2l+1)}{k^2}\sin^2\zeta_l = \frac{4\pi(2l+1)}{k^2}\cdot\left[\frac{\tan^2\zeta_l}{1+\tan^2\zeta_l}\right] \qquad (7.87^1)$$

Let $E_0^{(l)}$, be the energy at which $\beta_l = \Delta_l$. For E in the neighbourhood of $E_0^{(l)}$, we may write,

$$\beta_l(E) \approx \beta_l(E_0^{(l)}) + (E - E_0^{(l)})\left(\frac{\partial\beta_l}{E}\right)^{E=E_0^{(l)}}$$

$$\approx \Delta_l - b_l(E - E_0^{(l)}), \qquad (7.88)$$

where, $b_l > 0$ (see Footnote 21)[22]. Substituting in (7.85[1]) from (7.88), we get,

$$\tan\zeta_l = \frac{\Gamma_l}{2(E_0^{(l)} - E)}, \qquad (7.85^2)$$

where

$$\Gamma_l = 2(s_l/b_l). \qquad (7.89)$$

Substituting for $\tan\zeta_l$ from (7.85[2]) in (7.87[1]), we have,

$$\sigma^{(l)}(E) \approx \frac{4\pi(2l+1)}{k^2}\left[\frac{\Gamma_l^2}{4(E - E_0^{(l)})^2 + \Gamma_l^2}\right]. \qquad (7.87^2)$$

In Fig. 7.8, we show the variation of $\sigma_l(E)$ with E for reasonably small values of Γ_l. The cross-section has a sharp maximum centred around $E = E_0^{(l)}$. From Eq. (7.87[2]), we have,

---

21. By a procedure identical to the one leading to Eq. (7.69), we can show that

$$\beta_l(E_1) - \beta_l(E_2) = \frac{-(2\mu/\hbar^2)(E_1 - E_2)\int_0^{r_0} u_{k_1,l}(r)\,u_{k_2,l}(r)\,r_0\,dr}{u_{k_1,l}(r_0)\,u_{k_2,l}(r_0)},$$

showing that β_l is a *decreasing* function of E. The energy dependence of Δ_l is evident from Eq. (7.83b).

22. Since for the low energy, the wavefunction in the region $r < r_0$ is determined by the potential rather than the energy [Eqs. (7.59a) and (7.60)], the assumption that β_l is linear in E is a good approximation.

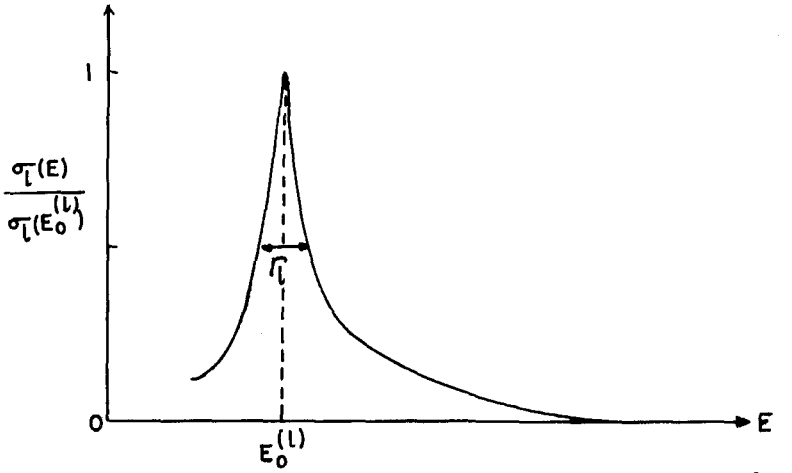


Fig. 7.8 Variation of the partial cross-section with energy near resonance [Eq. (7.87[2])].

$$\sigma^{(l)}(E_0^{(l)}) = \frac{4\pi(2l+1)}{k_0^2} = 4\pi(2l+1)(\hbar^2/2\mu E_0^{(l)}). \qquad (7.87^3)$$

The meaning of Γ_l should be clear from Eq. (7.87[2]) and Fig. 7.8. It is the *width* of the resonance peak. That is, $(\Gamma_l/2)$ is the value of $|E - E_0^{(l)}|$ for which $\sigma^{(l)}$ falls to half its peak value. Eq. (7.87[2]) is known as the (single-level) *Breit-Wigner formula.*

Now, in the case of *s*-wave scattering, we have seen (Fig. 7.6) that the existence of a bound state is implied whenever δ_0 crosses an odd multiple of $\pi/2$. We see from Eqs. (7.85[1]) that at resonance ($\beta_l \approx \Delta_l$), ζ_l equals an odd multiple of $\pi/2$ (so that $\delta_l \gtrsim (2n+1)\pi/2$, with $n = 0, 1, 2, \ldots$). Thus, scattering resonance for the lth partial wave at the energy $E_0^{(l)}$ is related to the existence of a bound state of angular momentum l near this energy. There is, however, a slight difference between the resonance energy $E_0^{(l)}$ and the energy $E_b^{(l)}$ of a true bound state. Whereas $E_b^{(l)}$ is negative $E_0^{(l)}$ is positive (albeit small). Thus, the resonance state is not a true bound state. For this reason, it is called a *virtual*, or a *metastable*, state. Its existence could be understood on the basis of the effective potential, composed of the actual potential plus the "centrifugal potential" $l(l+1)\hbar^2/2\mu r^2$ (the curve V_{eff} in Fig. 7.3). For a particle with near-zero energy that coincides with an energy level of V_{eff}, the latter acts as a *potential barrier* that slows down its escape from the potential well. The particle is so to say, "captured" by the potential and later re-emitted as a scattered particle. For a given energy E of the particle and for a given potential $V(r)$, it is the l-value that decides whether V_{eff} constitutes such a potential barrier or not (see Fig. 7.9)[23]. Usually, the potential $V(r)$ has a true

23. The quantum mechanical penetration of a potential barrier is discussed in Section 8.1.

bound state of angular momentum l just below zero energy when V_{eff} has a (virtual) bound state of the same angular momentum just above zero energy.

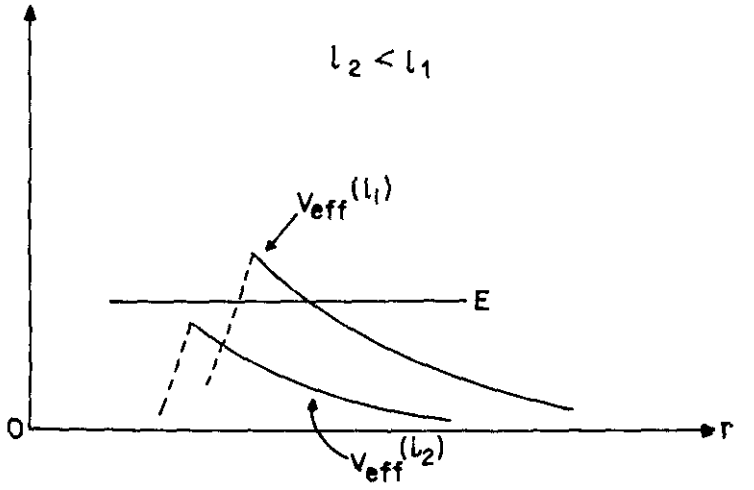


Fig. 7.9 Variation of the effective potential [Eq. (7.47)] with angular momentum.

The preceding discussion provides us with the following picture of the scattering phenomenon in the low-energy region: When the energy is far from a resonance energy, the particles do not penetrate into the interior region of the potential. It is as if the particles have met a hard sphere of radius 'a' (= the scattering length). Very few of the particles in the incident beam are scattered, and the scattering is
'instantaneous'. In the neighbourhood of a resonance energy, particles (with the appropriate angular momentum) begin to penetrate deep into the interior of the scattering region. They are trapped by the potential, but eventually are re-emitted as scattered particles. The delay between the time the particles enter the potential and the time they are re-emitted, is of the order of $\hbar/\Gamma_l$.

Problem 7.5: Derive the relationships (7.81), (7.82 a, b) and (7.83 b, c).

7.3 THE BORN APPROXIMATION

The usefulness of the method of partial waves, discussed in the previous section, is limited to the case of low-energy scattering by short-range central potentials. We will now discuss an approximation method which is suitable for large energies and where the potential is not necessarily central.

Substituting from Eq. (7.5), equation (7.4) of the scattering problem becomes,

$$(-\Delta - k^2)\psi(\mathbf{r}) = -U(\mathbf{r})\psi(\mathbf{r}), \tag{7.90}$$

where, $\hat{U}(\mathbf{r}) = (2\mu/\hbar^2)\hat{V}(\mathbf{r})$. This differential equation can be converted to an integral equation with the help of *Green's functions*. The technique is a common one applicable to any inhomogeneous differential equation of the form:

$$(\Omega - \omega_0)v(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \tag{7.91}$$

where $\hat{\Omega}$ is a Hermitian differential operator and $\rho(\mathbf{r})$ is the *inhomogeneity* (also referred to as *source density*). The solution of Eq. (7.91) will consist of a *particular integral* $w_{\omega_0}(\mathbf{r})$, plus a *complementary function* $u_{\omega_0}(\mathbf{r})$. The latter is a solution of the homogeneous equation,

$$(\hat{\Omega} - \omega_0)\, u_{\omega_0}(\mathbf{r}) = 0.$$

That is,

$$\Omega u_{\omega_0}(\mathbf{r}) = \omega_0 u_{\omega_0}(\mathbf{r}). \tag{7.92a}$$

Thus, $u_{\omega_0}(\mathbf{r})$ is an eigenvector of the operator $\hat{\Omega}$ belonging to the eigenvalue ω_0.

Now, $\hat{\Omega}$ has, in general, several eigenvalues and eigenvectors. Let us denote an arbitrary eigenvector by $u_\omega(\mathbf{r})$ and the set of all eigenvectors by $\{u_\omega(\mathbf{r})\}$. Since $\hat{\Omega}$ is Hermitian, the set of eigenvectors is a complete orthonormal one. That is[24] [see Footnote 6, Chapter 2].

$$\int u_{\omega'}^*(\mathbf{r}) u_\omega(\mathbf{r}) d^3\mathbf{r} = \delta(\omega - \omega'), \tag{7.93a}$$

$$\int u_\omega(\mathbf{r}) u_\omega^*(\mathbf{r}') d\omega = \delta(\mathbf{r} - \mathbf{r}'). \tag{7.93b}$$

Also, by definition,

$$\hat{\Omega} u_\omega(\mathbf{r}) = \omega u_\omega(\mathbf{r}). \tag{7.92b}$$

The particular integral $w_{\omega_0}(\mathbf{r})$ can be expressed in terms of the Green's function for the operator $(\hat{\Omega} - \omega_0\hat{1})$. The latter is defined by

$$(\Omega - \omega_0) G_{\omega_0}(\mathbf{r}, \mathbf{r}') = 4\pi\delta(\mathbf{r} - \mathbf{r}'). \tag{7.94}$$

Thus, the Green's function $G_{\omega_0}(\mathbf{r}, \mathbf{r}')$ is a solution of the homogeneous differential equation (7.92a), except at the point $\mathbf{r} = \mathbf{r}'$. This fact suggests the form:

$$G_{\omega_0}(\mathbf{r}, \mathbf{r}') = \int u_\omega(\mathbf{r}) f_\omega(\mathbf{r}') d\omega. \tag{7.95^1}$$

Substituting this expression in (7.94), we have,

$$\frac{1}{4\pi}\int (\omega - \omega_0) u_\omega(\mathbf{r}) f_\omega(\mathbf{r}') d\omega = \delta(\mathbf{r} - \mathbf{r}'). \tag{7.94a}$$

Comparing Eqs. (7.94a) and (7.93b), we get,

$$f_\omega(\mathbf{r}') = 4\pi\left\{\frac{u_\omega^*(\mathbf{r}')}{\omega - \omega_0}\right\},$$

and

$$G_{\omega_0}(\mathbf{r}, \mathbf{r}') = 4\pi \int \frac{u_\omega(\mathbf{r}) u_\omega^*(\mathbf{r}')}{\omega - \omega_0} d\omega. \tag{7.95^2}$$

---

24. We assume, for the sake of simplicity, that the eigenvalues are continuous.

Multiplying Eq. (7.94) with $\rho(\mathbf{r}')$ and integrating over $\mathbf{r}'$, we get[25],

$$(\hat{\Omega}-\omega_0)\int G_{\omega_0}(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')d^3\mathbf{r}' = 4\pi\int \rho(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')d^3\mathbf{r}' = 4\pi\rho(\mathbf{r}),$$

or,

$$w_{\omega_0}(\mathbf{r}) \equiv \frac{-4\pi\rho(\mathbf{r})}{\hat{\Omega}-\omega_0} = -\int G_{\omega_0}(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')d^3\mathbf{r}'. \tag{7.96}$$

The general solution of Eq. (7.91) is, thus, given by,

$$v_{\omega_0}(\mathbf{r}) = u_{\omega_0}(\mathbf{r}) + w_{\omega_0}(\mathbf{r}) \tag{7.97}$$

$$= u_{\omega_0}(\mathbf{r}) - 4\pi\int\int \frac{u_\omega(\mathbf{r})u_\omega^*(\mathbf{r}')}{\omega-\omega_0}\rho(\mathbf{r}')d\omega d^3\mathbf{r}'.$$

The integrand in Eq. (7.96) could be interpreted as the contribution to the particular integral that has its source in the volume element $d^3\mathbf{r}'$, so that $w_{\omega_0}(\mathbf{r})$ is the superposition of the contributions from all such volume elements where the source density is non-zero.

Applying the above procedure to Eq. (7.90), we get,

$$\psi_k(\mathbf{r}) = u_k(\mathbf{r}) - \tfrac{1}{4\pi}\int G_k(\mathbf{r},\mathbf{r}')U(\mathbf{r}')\psi_k(\mathbf{r}')d^3\mathbf{r}'. \tag{7.98}$$

with,

$$u_k(\mathbf{r}) = \frac{1}{(2\pi)^{3/2}}\exp(i\mathbf{k}\cdot\mathbf{r}), \tag{7.99}$$

and[27]

$$G_k(\mathbf{r},\mathbf{r}') = 4\pi\int \frac{u_k'(\mathbf{r})u_k^*(\mathbf{r}')}{(k')^2-k^2}d^3\mathbf{k}'$$

$$= \frac{1}{2\pi^2}\int \frac{\exp[i\mathbf{k}'\cdot(\mathbf{r}-\mathbf{r}')]}{(k')^2-k^2}d^3\mathbf{k}', \tag{7.100}$$

The integrand in Eq. (7.100) is a scalar so that the integral is independent of the co-ordinate system in which it is evaluated. Choosing, then, a co-ordinate system in which the z-axis is along the vector $\rho = \mathbf{r}-\mathbf{r}'$, (see Fig. 7.10), we have, $d^3\mathbf{k}' = k'^2dk'\sin\theta\, d\theta\, d\phi$ and,

$$G_k(\mathbf{r},\mathbf{r}') = (2\pi i\rho)^{-1}\int_{-\infty}^{+\infty}\frac{\kappa(e^{i\kappa}-e^{-i\kappa})}{\kappa^2-\sigma^2}d\kappa, \tag{7.100a}$$

where $\kappa = k'\rho$; $\sigma = k\rho$.

The integral in (7.100a) can be evaluated with the help of the *theorem of residues*[27] which states: "If $F(z)$ is a function of the complex variable z such that it is

25. $\hat{\Omega}$ operates only on the co-ordinate $\mathbf{r}$, so that it can be taken outside the integral sign.

26. Note that the Green's function is characterised only by the magnitude of $\mathbf{k}$ (that is, by k^2) rather than by the vector $\mathbf{k}$.

27. See, Arfken, G. *Mathematical methods for physicists* (Academic Press, New York, 1970) II edition, Section 7.2.

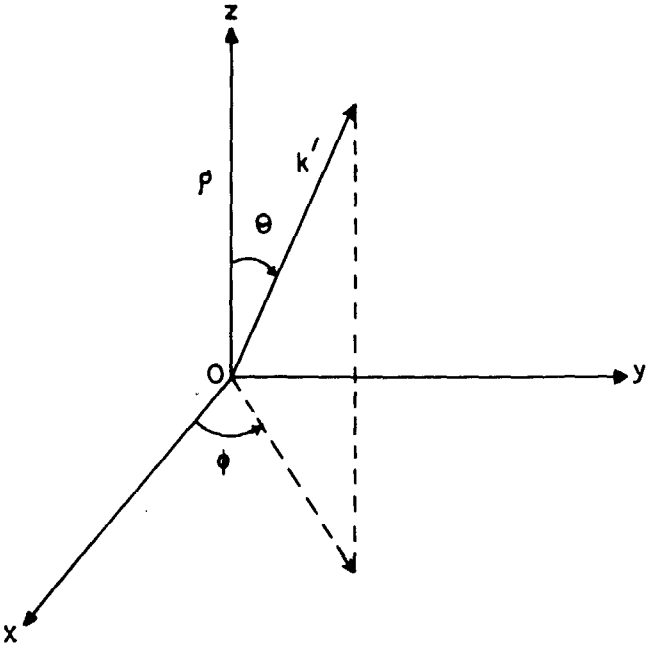


Fig. 7.10

analytic throughout a (closed) contour C and its interior, except at a number of poles[28] *inside* the contour, then,

$$\oint_C F(z)dz = 2\pi i \Sigma R, \tag{7.101}$$

where, ΣR denotes the sum of the residues of $F(z)$ at those of its poles that are situated within the contour''. The arrow on the circle on the integral sign denotes the positive sense of integration. The residue at the (simple) pole $z = a$ is given by

$$R(at\ a) = \underset{z \to a}{\text{Lt}} \{(z-a)F(z)\}. \tag{7.102}$$

According to *Jordan's lemma*[27], the integral in Eq. (7.100a) can be written as

$$I(\sigma) = I_1(\sigma) + I_2(\sigma), \tag{7.103}$$

with

$$I_1(\sigma) = \oint_{C_1} \left(\frac{ze^{iz}}{z^2 - \sigma^2} \right) dz \equiv \oint_{C_1} F_1(z)dz, \tag{7.104a}$$

---

28. The point $z = a$ is a *non-essential singularity* of $F(z)$ if $F(a)$ is infinite but $\{(z-a)^m F(z)\}_{z=a} \equiv F_m(a)$, is finite, where m is a positive integer. That is, a non-essential singularity at 'a' is removable by multiplying the function by the mth power of $(z-a)$. If the singularity is not thus removable with any finite value of m, then the singularity is *essential*. A pole is a non-essential singularity and the smallest value of m for which $F_m(a)$ is finite, is the *order* of the pole. When $m = 1$, we have a *simple* pole.

and

$$I_2(\sigma) = -\oint_{C_2}\left(\frac{ze^{-iz}}{z^2-\sigma^2}\right)dz \equiv -\oint_{C_2} F_2(z)dz, \qquad (7.104b)$$

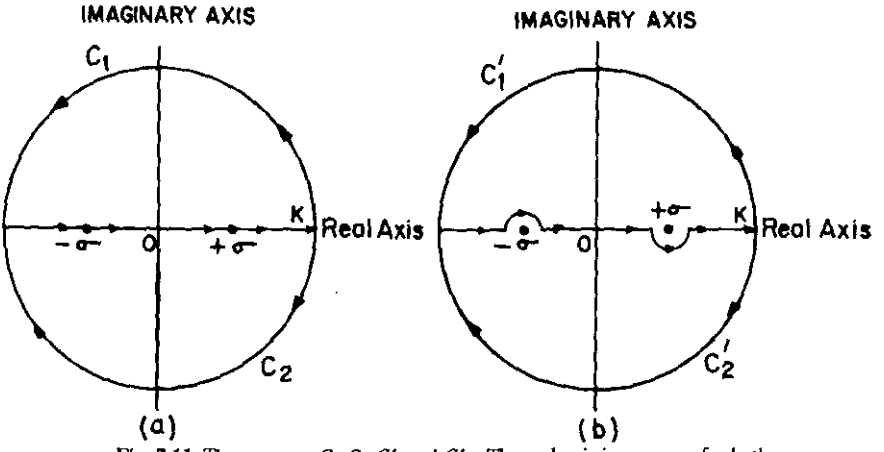


Fig. 7.11 The contours C_1, C_2, C'_1 and C'_2. The real axis is common for both lower and the upper contours.

where C_1 consists of the real axis and an infinite semi-circle in the upper half-plane while C_2 consists of the real axis and an infinite semi-circle in the lower half-plane [Fig. 7.11(a)]. However, since the poles $z = \pm\sigma$ are *on* the contour (so that $F(z)$ is not analytic *throughout* the contour), a straight forward application of the residue theorem is not possible. In fact, the integrals $I_1(\sigma)$ and $I_2(\sigma)$ are *improper* for this reason. Just as a nonanalytic function has no unique derivative (for, the derivative depends on the direction from which we approach the point of interest as, for example, in the case of the derivative of the function $F(x)$ at $x = 0$ in Fig. C.1), the value of an improper integral depends on the limiting process used to evaluate the integral. In the present case, the possible limiting processes are:

(i) Deform the contour to C'_1 and C'_2 as shown in Fig. 7.11(b).[29] Then,

$$I_1(\sigma) = \operatorname*{Lt}_{r\to 0}\oint_{C'_1} F_1(z)dz,$$

and

$$-I_2(\sigma) = \operatorname*{Lt}_{r\to 0}\oint_{C'_2} F_2(z)dz.$$

This leads to[30]

$$I(\sigma) = 2\pi i \cos\sigma, \qquad (7.103^1)$$

29. Actually, there are three other ways of deforming the contour. Thus, we can go under the real axis at the pole $\kappa = -\sigma$ and over the real axis at $\kappa = +\sigma$, or go either over or under the real axis at both places. All these lead to the same result, Eq. (7.103^1).

30. See, Example 7.2.3 in the reference quoted in footnote 27.

and

$$G_k(\mathbf{r},\mathbf{r}') = \frac{\cos(k\,|\,\mathbf{r}-\mathbf{r}'\,|)}{|\,\mathbf{r}-\mathbf{r}'\,|} \qquad (7.100^1)$$

(ii) Replace σ by $\sigma + i\eta$, where $\eta > 0$ but is small. Then,

$$I(\sigma) = \underset{\eta\to 0}{\text{Lt}}\, I(\sigma + i\eta).$$

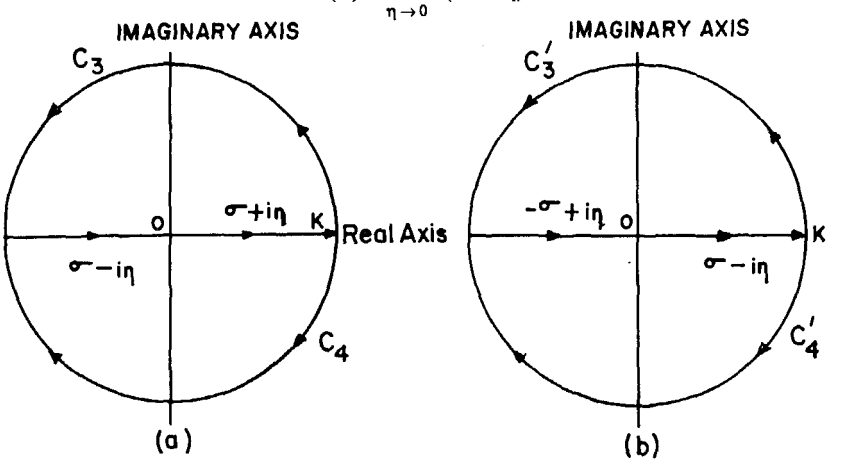


Fig. 7.12 Contours for the integrals leading to Eqs. (7.103²) and (7.103³).

This procedure shifts the pole away from the real axis and thus away from the contour [Fig. 7.12(a)]. After integration, the poles are brought back. Application of the residue theorem gives,

$$I_1(\sigma+i\eta) \equiv \oint_{C_3} \frac{ze^{iz}}{(z-\sigma-i\eta)(z+\sigma+i\eta)}\,dz$$

$$= 2\pi i \left\{\frac{ze^{iz}}{z+\sigma+i\eta}\right\}_{z=\sigma+i\eta}$$

$$= \pi i e^{i(\sigma+i\eta)},$$

and

$$I_2(\sigma+i\eta) = +(2\pi i)\left\{\frac{ze^{-iz}}{z-\sigma-i\eta}\right\}_{z=-\sigma-i\eta} = \pi i e^{i(\sigma+i\eta)},$$

so that,

$$I(\sigma) = 2\pi i e^{i\sigma}, \qquad (7.103^2)$$

$$G_k(\mathbf{r},\mathbf{r}') = \frac{e^{ik\,|\,\mathbf{r}-\mathbf{r}'\,|}}{|\,\mathbf{r}-\mathbf{r}'\,|}. \qquad (7.100^2)$$

(iii) Replace σ by $\sigma - i\eta$ [Fig. 7.12(b)]. This leads to

$$I(\sigma) = \underset{\eta\to 0}{\text{Lt}}\, I(\sigma - i\eta) = 2\pi i e^{-i\sigma}, \qquad (7.103^3)$$

and

$$G_k(\mathbf{r},\mathbf{r}') = \frac{e^{-ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}. \qquad (7.100^3)$$

Eqs. ($7.100^{1,2,3}$) represent, respectively, a standing spherical wave, an outgoing spherical wave and an incoming spherical wave. From the physics point of view, choosing the limiting process is, thus, equivalent to choosing the boundary conditions. In the present case, the boundary conditions require an outgoing spherical wave in (7.98) [see Eq. (7.8)]. Therefore, the limiting process to be used is the one described in (ii) above, and the Green's function is given by Eq. (7.100^2).

Now, for $r \gg r'$,

$$\begin{aligned} |\mathbf{r}-\mathbf{r}'| &= (r^2 + r'^2 - 2rr'\cos\theta)^{1/2} \\ &\approx r\{1 - 2(r'/r)\cos\theta\}^{1/2} \\ &\approx r - r'\cos\theta, \end{aligned}$$

where θ is the angle between $\mathbf{r}$ and $\mathbf{r}'$.

Then,

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} \approx \frac{1}{r},$$

so that,

$$G_k(\mathbf{r},\mathbf{r}') \underset{r\gg r'}{\sim} \frac{e^{ikr}}{r} e^{-i\mathbf{k}'\cdot\mathbf{r}'}. \qquad (7.104)$$

Substituting this value of $G_k(\mathbf{r},\mathbf{r}')$ in (7.98) and neglecting the normalization factor in $u_k(\mathbf{r})$, we have,

$$\psi_k(\mathbf{r}) \underset{r\to\infty}{\sim} \exp(i\mathbf{k}.\mathbf{r}) - \frac{1}{4\pi}\frac{e^{ikr}}{r}\int \exp(-i\mathbf{k}'.\mathbf{r}')U(\mathbf{r}')\psi_k(\mathbf{r}')d^3\mathbf{r}'. \qquad (7.105)$$

Comparing Eq. (7.105) with Eq. (7.8), we have,

$$f_k(\theta,\phi) = -\frac{1}{4\pi}\int \exp(-i\mathbf{k}'\cdot\mathbf{r}')U(\mathbf{r}')\psi_k(\mathbf{r}')d^3\mathbf{r}' \qquad (7.106)$$

Integral Equation for Scattering

Since ψ_k occurs on both sides, Eq. (7.105) does not really represent a solution of the differential equation (7.90); it is actually the *integral equation* for scattering. That is, the Green's function has helped us only to convert the differential equation to an integral equation. But now the advantage is that we can solve the equation by an approximation method applicable to an integral equation (but not applicable to a differential equation), namely, the *method of iteration*. This is a successive approximation method in which the nth approximation to $\psi_k(\mathbf{r})$ on the

left hand side of Eq. (7.98) is obtained by substituting the $(n-1)$th approximation to $\psi_{\mathbf{k}}(\mathbf{r}')$ on the right hand side of that equation. Thus, writing the equation in terms of Green's functions, we have,[31]

$$\psi_{\mathbf{k}}^{(n)}(\mathbf{r}) \approx \exp(i\mathbf{k}\cdot\mathbf{r}) - \frac{1}{4\pi}\int G_k(\mathbf{r},\mathbf{r}_1)U(\mathbf{r}_1)\psi_{\mathbf{k}}^{(n-1)}(\mathbf{r}_1)d^3\mathbf{r}_1, \qquad (7.107^1)$$

where $\psi_{\mathbf{k}}^{(n)}$ is $\psi_{\mathbf{k}}$ to the nth approximation.

Writing,

$$\psi_{\mathbf{k}}^{(n-1)}(\mathbf{r}_1) = \exp(i\mathbf{k}\cdot\mathbf{r}_1) - \frac{1}{4\pi}\int G_k(\mathbf{r}_1,\mathbf{r}_2)U(\mathbf{r}_2)\psi_{\mathbf{k}}^{(n-2)}(\mathbf{r}_2)d^3\mathbf{r}_2,$$

$$\psi_{\mathbf{k}}^{(n-2)}(\mathbf{r}_2) = \exp(i\mathbf{k}\cdot\mathbf{r}_2) - \frac{1}{4\pi}\int G_k(\mathbf{r}_2,\mathbf{r}_3)U(\mathbf{r}_3)\psi_{\mathbf{k}}^{(n-3)}(\mathbf{r}_3)d^3\mathbf{r}_3,$$

and so on, we get,

$$\psi_{\mathbf{k}}^{(n)}(\mathbf{r}) = \sum_{p=0}^{n}\Phi_{\mathbf{k}}^{p}(\mathbf{r}), \qquad (7.107^2)$$

with

$$\Phi_{\mathbf{k}}^{p}(\mathbf{r}) = \left(\frac{-1}{4\pi}\right)^p \int G_k(\mathbf{r}_0,\mathbf{r}_1)U(\mathbf{r}_1)G_k(\mathbf{r}_1,\mathbf{r}_2)U(\mathbf{r}_2)$$

$$\ldots G_k(\mathbf{r}_{p-1},\mathbf{r}_p)U(\mathbf{r}_p)\exp(i\mathbf{k}\cdot\mathbf{r}_p)d^3\mathbf{r}_1\ldots d^3\mathbf{r}_p, \qquad (7.108)$$

where $\mathbf{r}_0 \equiv \mathbf{r}$.

In order to understand the meaning of $\Phi_{\mathbf{k}}^{p}$, let us look more closely at $\Phi_{\mathbf{k}}^{2}$. We have,

$$\Phi_{\mathbf{k}}^{2}(\mathbf{r}) = \left(-\frac{1}{4\pi}\right)^2 \int\int G_k(\mathbf{r},\mathbf{r}_1)U(\mathbf{r}_1)G_k(\mathbf{r}_1,\mathbf{r}_2)U(\mathbf{r}_2).\exp(i\mathbf{k}\cdot\mathbf{r}_2)d^3\mathbf{r}_1 d^3\mathbf{r}_2$$

$$= \left(\frac{-1}{4\pi}\right)^2 \int d^3\mathbf{r}_1 G_k(\mathbf{r},\mathbf{r}_1)U(\mathbf{r}_1)\int d^3\mathbf{r}_2 G_k(\mathbf{r}_1,\mathbf{r}_2)U(\mathbf{r}_2)\exp(i\mathbf{k}\cdot\mathbf{r}_2). \qquad (7.108a)$$

This has the following obvious interpretation: A plane wave $\exp(i\mathbf{k}\cdot\mathbf{r}_2)$ is incident at the volume element $d^3\mathbf{r}_2$. It gets scattered there by the potential $U(\mathbf{r}_2)$. The scattered wave is then propagated from $d^3\mathbf{r}_2$ to the volume element $d^3\mathbf{r}_1$, this propagation being represented by the Green's function[32] $G_k(\mathbf{r}_1\mathbf{r}_2)$. At $d^3\mathbf{r}_1$, the wave is scattered again by the potential $U(\mathbf{r}_1)$. Finally, the wave is propagated from $d^3\mathbf{r}_1$ to the point of observation $\mathbf{r}$ by $G_k(\mathbf{r},\mathbf{r}_1)$. The whole process is schematically shown in Fig. (7.13). Thus, $\Phi_{\mathbf{k}}^{2}(\mathbf{r})$ represents the contribution to $\psi_{\mathbf{k}}^{(n)}(\mathbf{r})$ from all such doubly-scattered waves, and $\Phi_{\mathbf{k}}^{p}(\mathbf{r})$ is the contribution from waves (particles) that have been scattered p times by the potential. The maximum number of scattering suffered by a particle contained in $\psi_{\mathbf{k}}^{(n)}(\mathbf{r})$ is, thus, n, and

---

31. Eq. (7.107^1) is known as the *n*th *iterated form* of Eq. (7.98).
32. The Green's function is, for this reason, called the *propagator*.

$$\Phi_k^0(\mathbf{r}) \equiv \psi_k^{(0)}(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}),$$

represents particles that have suffered no scattering at all (that is, the incident wave).

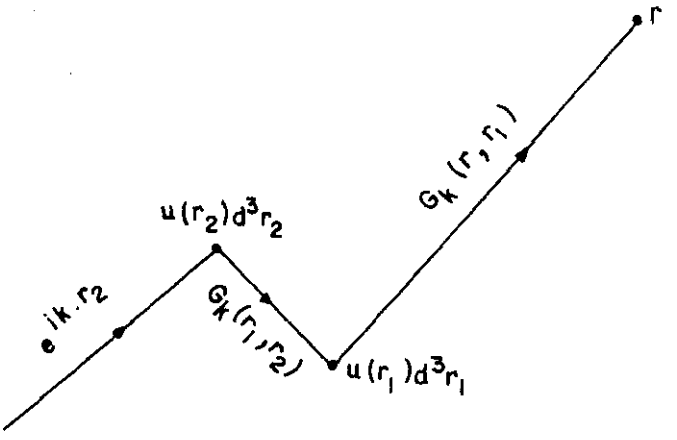


Fig. 7.13 Schematic representation of the doubly scattered particles reaching the point of observation $\mathbf{r}$.

The series (7.107[2]) in the limit $n \to \infty$ is known as the *Neumann series*. It represents a solution to the scattering problem [Eq. (7.90) or Eq. (7.98)] if the series is a converging one. Roughly speaking, such a condition would be satisfied if the wave gets weaker and weaker at successive scatterings. That is, if the number of particles getting scattered progressively diminishes at each successive scattering. We see from Eq. (7.108) that a necessary condition for this is that the potential be weak.[33]

The approximation,

$$\psi_k(\mathbf{r}) \approx \psi_k^{(n)}(\mathbf{r}), \tag{7.105a}$$

is known as the nth *Born Approximation*.

The scattering amplitude in the nth Born Approximation is given, according to Eq. (7.106), by

$$f_k^{(n)}(\theta, \phi) = -\frac{1}{4\pi}\int \exp(-i\mathbf{k}'\cdot\mathbf{r}')U(\mathbf{r}')\psi_k^{(n-1)}(\mathbf{r}')d^3\mathbf{r}'. \tag{7.106a}$$

If the potential is weak enough (so that the convergence of the Neumann series is rapid enough), the *first Born Approximation* (which is also called simply *the* Born Approximation) provides a good enough approximation to $\psi_k(\mathbf{r})$:

$$\psi_k(\mathbf{r}) \approx \psi_k^{(1)}(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r}) - \frac{1}{4\pi}\int G_k(\mathbf{r}, \mathbf{r}')U(\mathbf{r}')\exp(i\mathbf{k}\cdot\mathbf{r}')d^3\mathbf{r}'$$

$$\underset{r\to\infty}{\sim} \exp(i\mathbf{k}\cdot\mathbf{r}) - \frac{1}{4\pi}\frac{e^{ikr}}{r}\int \exp(i\mathbf{K}\cdot\mathbf{r}')U(\mathbf{r}')d^3\mathbf{r}', \tag{7.105b}$$

33. A more exact criterion will be derived later in the case of the first Born Approximation [Eq. (7.112b)].

and

$$f_k(\theta,\phi) \approx f_k^{(1)}(\theta,\phi) = -\frac{1}{4\pi}\int \exp(i\mathbf{K}\cdot\mathbf{r}')U(\mathbf{r}')d^3\mathbf{r}', \qquad (7.106b)$$

where $\hbar\mathbf{K} = \hbar(\mathbf{k}-\mathbf{k}')$, is the momentum transferred from the particle to the potential. Referring to Fig. (7.14), we have,

$$K \equiv |\mathbf{K}| = (k^2 + k'^2 - 2kk'\cos\theta)^{1/2}$$

$$= \{2k^2(1-\cos\theta)\}^{1/2}$$

$$= 2k\sin(\theta/2), \qquad (7.109)$$

since

$$|\mathbf{k}| = |\mathbf{k}'|$$

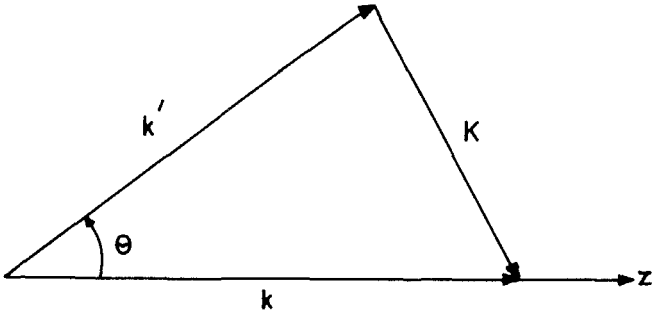


Fig 7.14

Eq. (7.106b) shows that the scattering amplitude is just proportional to the Fourier transform of the potential in the 'momentum-transfer space.'[34]

When the potential is central $U(\mathbf{r}) \equiv U(r)$, we can further simplify the expression (7.106b) by taking advantage of the fact that the integrand is a scalar. Thus, choosing a co-ordinate system in which the z-axis is along $\mathbf{K}$, we get,

$$f_k(\theta,\phi) \equiv f_k(\theta) = -\frac{1}{K}\int_0^\infty rU(\mathbf{r})\sin Kr\,dr, \qquad (7.110)$$

which is the (first) Born Approximation scattering amplitude in the case of a central potential. We note the following important features:

(i) The amplitude is independent of the azimuthal angle ϕ. This is a consequence of the cylindrical symmetry of the potential and is in agreement with the result (7.35) obtained in the method of partial waves.

(ii) It depends only on the momentum transfer (which is proportional to $k\sin(\theta/2)$) and not on the momentum of the incident particle or on the scattering angle individually.

34. It is also proportional to the matrix element of the potential between the plane wave states $\exp(i\mathbf{k}\cdot\mathbf{r}')$ (the *initial* state) and $\exp(i\mathbf{k}'\cdot\mathbf{r}')$ (the *final* state).

When the potential is weak, the method of partial waves also can be shown to yield expression (7.110) for the scattering amplitude. In this case, the phase shift δ_l is given by Eq. (7.45[2]). Then from Eqs. (7.36) and (7.36a), we have,

$$f_k(\theta) = \frac{1}{k}\sum_{l=0}^{\infty}(2l+1)(1+i\delta_l+\ldots)\delta_l P_l(\cos\theta)$$

$$\approx \sum_{l=0}^{\infty}(2l+1)P_l(\cos\theta)\cdot(\delta_l/k)$$

$$\approx -\sum_{l=0}^{\infty}(2l+1)P_l(\cos\theta)\int_0^{\infty}[j_l(kr)]^2U(r)r^2dr.$$

$$\approx -\frac{1}{K}\int_0^{\infty} rU(r)\sin Kr\,dr, \qquad (7.110a)$$

where the addition theorem (E.75) for spherical Bessel functions has been used. Now, the addition theorem is valid only if a large number of terms contribute to the summation in (E.75). This requires that δ_l be nonzero for a large number of partial waves, which in turn requires that the energy of the incident particles be large. Thus, derivation (7.110a) is valid only when the potential is weak (so that δ_l is small) and the energy is large (so that δ_l is not zero). As we have seen, these are not conditions under which the method of partial waves can be usefully applied.

Criterion for the Validity of the Born Approximation

The criterion is that $|\Phi_k^1| \ll |\Phi_k^0| = 1$. Now, in the case of a central potential, we have, from Eqs. (7.100[2]) and (7.105b),

$$\Phi_k^1(\mathbf{r}) = -\frac{1}{4\pi}\int\frac{\exp(ik|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}U(r')\exp(i\mathbf{k}\cdot\mathbf{r}')d^3\mathbf{r}'.$$

It is reasonable to assume that Φ_k^1 has its maximum value at $r = 0$ (the centre of the potential). Now,

$$\Phi_k^1(0) = -\frac{1}{4\pi}\left(\frac{2\mu}{\hbar^2}\right)\int\frac{\exp(ikr')}{r'}V(r')\exp(i\mathbf{k}\cdot\mathbf{r}')d^3\mathbf{r}'.$$

$$= -\left(\frac{2\mu}{\hbar^2 k}\right)\int_0^{\infty}\exp(ikr')\sin kr'V(r')dr'$$

Let us assume that the potential is of finite range, say r_0, and of strength V_0. Then,

$$|\Phi_k^1(0)| = \frac{\mu V_0}{2\hbar^2k^2}|\exp(2ikr_0) - 2ikr_0 - 1|. \qquad (7.111)$$

Case 1: $kr_0 \ll 1$ (*low energy*)

In this case, expanding e^{2ikr_0} in terms of $2kr_0$, we get,

$$|\Phi_k^1(0)| \approx \frac{\mu V_0 r_0^2}{\hbar^2}.$$

Thus, the Born approximation would be applicable if

$$\frac{\mu V_0 r_0^2}{\hbar^2} \ll 1,$$

or,

$$V_0 r_0^2 \ll \frac{\hbar^2}{\mu}, \tag{7.112a}$$

But, if the potential is strong enough to cause binding, then[35]

$$V_0 r_0^2 \geq \left(\frac{\pi^2}{8}\right)\frac{\hbar^2}{\mu}, \tag{7.113}$$

so that, Born approximation will not be valid at low energies.

Case 2: $kr_0 \gg 1$

In this case,

$$|e^{2ikr_0} - 2ikr_0 - 1| \approx 2kr_0,$$

so that,

$$|\Phi_k^1(0)| \approx \frac{\mu V_0 r_0}{\hbar^2 k} \ll 1.$$

or,

$$\frac{V_0 r_0}{\hbar v} \ll 1, \tag{7.112b}$$

where,

$$v = \frac{\hbar k}{\mu},$$

is the velocity of the particle.

Thus, the kinetic energy should be large compared with the potential energy. This result is consistent with the remarks regarding the validity of derivation (7.110a). Even though the above result is derived for the special case of a short-range central potential, it is of more general validity[36].

Scattering of Electrons by Atoms

As an application of the Born Approximation formula (7.110), let us consider the scattering of electrons by atoms. In this case, the potential is the screened coulomb potential,

---

35. See, Ref. 1, Section 15.
36. For a more detailed discussion on this point, see Wu, T.W. and Ohmura, T. *Quantum Theory of Scattering* (Prentice Hall, New Jersey 1962), Sections C.3 and C.4.

$$V(r) = (Ze^2/r)e^{-r/r_0}, \tag{7.114}$$

where, Z is the atomic number[37]. Substituting from (7.114) in (7.110), we get [remembering, $U(r) = (2\mu/\hbar^2)V(r)$],

$$f_k(\theta) = -(2\mu/\hbar^2)Ze^2\left(\frac{r_0^2}{K^2r_0^2+1}\right). \tag{7.115a}$$

For a pure coulomb force, $r_0 = \infty$, so that Eq. (7.115a) reduces to

$$f_k(\theta) = -(2\mu Ze^2/\hbar^2K^2), \tag{7.115b}$$

and the differential scattering cross-section is given by

$$\frac{d\sigma}{d\Omega} = |f_k(\theta)|^2 = \left(\frac{Ze^2}{2\mu v^2}\right)^2 \frac{1}{\sin^4(\theta/2)}, \tag{7.116}$$

v being the velocity of the particle. It so happens that formula (7.116), obtained here as an approximation, is in agreement with both the classical Rutherford formula and the exact quantum mechanical formula for Coulomb scattering[38].

Problem 7.6: Use the Optical Theorem [Eq. (7.39)] to show that the Born Approximation cannot be expected to give the correct differential scattering cross-section in the forward ($\theta = 0$) direction.

Problem 7.7: Using the Born Approximation formula (7.110) show that scattering by a square-well potential of depth V_0 and range r_0 has the following characteristics:

(a) Scattering is peaked in the forward direction,

(b) At large energies, the total scattering cross-section is inversely proportional to the energy.

REFERENCES

1. SCHIFF, L.I., *Quantum Mechanics*, III Edition (McGraw-Hill–Kogakusha, Tokyo, 1968), Sections 18 and 19.

2. MERZBACHER, E., *Quantum Mechanics*, II Edition (John Wiley, New York 1970), Chapter 11.

37. The pure Coulomb potential (Ze^2/r) corresponds to $r_0 = \infty$. The effect of a finite r_0 is to reduce the range of the potential to a value which is of the order of r_0. Thus, r_0 could be regarded as the range of the potential.

38. See, Ref. 2, Section 11.8.

CHAPTER 8

APPROXIMATION METHODS

An exact solution of the Schrödinger equation is an impractical proposition except for the simplest of potentials. In most cases of practical interest, one has to settle for an approximate solution. Thus, several methods of approximation have come to be devised for tackling various types of problems in quantum mechanics. These methods could be broadly divided into two categories: those for time-independent problems and those for time-dependent problems. The former refer to the methods applicable to the time-independent Schrödinger equation (4.18) and the latter to those applicable to the time-dependent Schrödinger equation (4.14). In this chapter, we propose to consider some of these approximation methods.

A. METHODS FOR TIME-INDEPENDENT PROBLEMS

In this category, we will discuss the WKB Approximation, the Variational Method and the Time-independent Perturbation Theory.

8.1 THE WKB APPROXIMATION

The Principle of the Method

This approximation method, named after *Wentzel*, *Kramers* and *Brillouin* who first introduced the method in quantum mechanics[1], is also known by the alternative names, *semi-classical approximation* and *phase-integral method.* The method is suited only to problems in one dimension or to problems that can be decomposed into one or more one-dimensional ones. The principle underlying the method is elucidated in the following:

Consider the classical equation,

$$H = \frac{\mathbf{p}^2}{2\mu} + V(\mathbf{r}), \tag{8.1}$$

---

1. Wentzel, G., *Z. Physik*, **38**, 518 (1926);
Kramers, H.A., *Z. Physik*, **39**, 828 (1926);
Brillouin, L., *Compt. Rend.*, **183**, 24 (1926).

where, H is the Hamiltonian, $\mathbf{p}$ the momentum and μ the mass of a particle. The potential $V(\mathbf{r})$ represents the external field which influences the motion of the particle. Eq. (8.1) can be converted into the equation of motion of the particle by substituting for H and $\mathbf{p}$ from Eqs. (1.14a, b). We get,

$$-\frac{\partial S}{\partial t}=\frac{(\nabla S)^2}{2\mu}+V(\mathbf{r}), \tag{8.2}$$

where, S is the action associated with the classical path of the particle [see Eq. (1.6)]. Eq. (8.2) is the (time-dependent) Hamilton-Jacobi equation of classical mechanics[2]. The corresponding equation of motion in quantum mechanics, is the time-dependent Schrödinger equation (4.14) with the Hamiltonian given by Eq. (8.1) where $\mathbf{p}$ and V are replaced by operators $\hat{\mathbf{p}}$ and $\hat{V}$ in accordance with Postulate IV of Chapter 3. In the co-ordinate representation, we have, $\hat{\mathbf{p}} \equiv -i\hbar\nabla$ [see Eq. (3.18[1])] and $\psi(t) \rightarrow \langle \mathbf{r} \mid \psi(t)\rangle = \psi(\mathbf{r}, t)$, so that Eq. (4.14) reads,

$$i\hbar\frac{\partial\psi(\mathbf{r}.t)}{\partial t}=\left[-\frac{\hbar^2}{2\mu}\nabla^2+V(\mathbf{r})\right]\psi(\mathbf{r},t). \tag{8.3}$$

The connection between Eqs. (8.2) and (8.3) would become clear if we substitute for $\psi(\mathbf{r}, t)$ in the latter from Eq. (4.15b), namely,

$$\psi(\mathbf{r},t) = \exp\{(i/\hbar)S(\mathbf{r},t)\} \tag{8.4}$$

we get,

$$\left(\frac{-\partial S}{\partial t}\right)\psi=\left[\frac{(\nabla S\cdot\nabla S)}{2\mu}-\frac{i\hbar}{2\mu}\nabla^2 S+V(\mathbf{r})\right]\psi, \tag{8.3a}$$

where, the identity, $\operatorname{div}(\mathbf{A}\phi) = \mathbf{A}\cdot\nabla\phi + \phi \operatorname{div}\mathbf{A}$, is used in evaluating $\nabla^2\psi$. Thus the equation satisfied by S is given by,

$$-\frac{\partial S}{\partial t}=\frac{(\nabla S)^2}{2\mu}+V(\mathbf{r})-\frac{i\hbar}{2\mu}\nabla^2 S. \tag{8.5}$$

Comparing Eqs. (8.5) and (8.2), we see that quantum mechanics should reduce to classical mechanics in the limit $\hbar \rightarrow 0$ [This is evident also from Eq. (1.10a)]. In other words, the finite value of $\hbar$ is responsible for the difference between classical and quantum mechanics. Now $\hbar$, being a universal constant, cannot be equal to zero. What is possible, and is in effect equivalent to $\hbar \rightarrow 0$, is that the term containing $\hbar$ in (8.5) can be negligible compared with the term containing $(\nabla S)^2$. This suggests that when the condition,

$$|(\nabla S)|^2 \gg \hbar\,|\nabla^2 S|, \tag{8.6a}$$

or,

$$|\mathbf{p}^2| \gg \hbar\,|(\nabla\cdot\mathbf{p})|, \tag{8.6b}$$

---

2. See, footnote 1, Chapter 4.

is satisfied, an approximation method based on a power series expansion of S in $\hbar$ is possible. Thus, writing[3],

$$S = S_0 + \frac{\hbar}{i} S_1 + \left(\frac{\hbar}{i}\right)^2 S_2 + \ldots, \tag{8.7}$$

the classical approximation consists in neglecting all powers of $\hbar$ higher than zero. In the semi-classical or WKB approximation, the terms up to the first power of $\hbar$ is retained. Thus, the WKB approximation is just a step ahead of the classical approximation.

The WKB Wavefuntion

As we stated earlier, practical applications of the WKB approximation is limited to time-independent (that is, stationary) problems in one dimension. We, therefore, confine our attention to such problems.

In the case of stationary problems, we have, from Eq. (4.21),

$$\psi(\mathbf{r}, t) = \Phi(\mathbf{r}) \exp\{-(i/\hbar)Et\}, \tag{8.8}$$

with [see Eq. (8.4)]

$$\Phi(\mathbf{r}) = \exp[(i/\hbar)W(\mathbf{r})], \tag{8.9}$$

where,

$$S(\mathbf{r}, t) = W(\mathbf{r}) - Et, \tag{8.10}$$

E being the energy of the system.

Thus, for the one-dimensional case, Eqs. (8.2), (8.5) and (8.7) reduce, respectively, to the equations,

$$\left(\frac{dW_0}{dx}\right)^2 - 2\mu[E - V(x)] = 0, \tag{8.2^1}$$

$$\left(\frac{dW}{dx}\right)^2 - 2\mu[E - V(x)] - i\hbar\frac{d^2W}{dx^2} = 0, \tag{8.5^1}$$

and

$$W(x) = W_0(x) + \left(\frac{\hbar}{i}\right)W_1(x) + \left(\frac{\hbar}{i}\right)^2 W_2(x) + \ldots, \tag{8.7^1}$$

while, the Schrödinger equation (8.3) reduces [with the help of Eq. (8.8)] to,

$$\frac{d^2\Phi}{dx^2} + \frac{2\mu}{\hbar^2}[E - V(x)]\Phi = 0, \tag{8.3^1}$$

with

---

3. Note that the dimensions of the S_k are such that $\hbar^{k-1}S_k$ is dimensionless. S is expanded in powers of $(\hbar/i)$ rather than of $\hbar$, because of the occurrence of the factor $(i/\hbar)$ in ψ [(Eq. (8.4)].

$$\Phi(x) = \exp\left[\frac{i}{\hbar}W(x)\right]. \tag{8.9^1}$$

From Eq. (8.2^1), we have,[4]

$$2\mu[E - V(x)] = \left(\frac{dW_0}{dx}\right)^2 = p^2(x), \text{ say}. \tag{8.11}$$

Then, Eq. (8.3^1) becomes,

$$\frac{d^2\Phi}{dx^2} + \frac{p^2}{\hbar^2}\Phi = 0. \tag{8.3^2}$$

We are interested in the solution of this equation within the WKB approximation. This is obtained by substituting for W in Eq. (8.9^1) the approximate value, $W \approx W_0 + (\hbar/i)W_1$ from the series expression (8.7^1). The values of W_0 and W_1 appropriate to the system being considered, are determined by substituting (8.7^1) in Eq. (8.5^1) which is equivalent to the Schrödinger equation (8.3^2). The substitution,

$$u(x) = \frac{dW(x)}{dx}, \tag{8.12}$$

will prove convenient. Substituting from Eqs. (8.11) and (8.12), Eq. (8.5^1) reduces to

$$\frac{\hbar}{i}\frac{du}{dx} = p^2 - u^2. \tag{8.5^2}$$

Also, differentiating Eq. (8.7^1) with respect to x, we get,

$$u(x) = u_0(x) + \left(\frac{\hbar}{i}\right)u_1(x) + \left(\frac{\hbar}{i}\right)^2 u_2(x) + \ldots, \tag{8.7^2}$$

where,

$$u_i(x) = \frac{dW_i}{dx}. \tag{8.13}$$

In terms of u, $\Phi(x)$ is given by,

$$\Phi(x) = \exp\left[\frac{i}{\hbar}\int^x \frac{dW}{dx'}dx'\right]$$

$$= \exp\left[\frac{i}{\hbar}\int^x u\, dx'\right]. \tag{8.9^2}$$

Substituting (8.7^2) in Eq. (8.5^2), we have,

$$\left(\frac{\hbar}{i}\right)\frac{du_0}{dx} + \left(\frac{\hbar}{i}\right)^2\frac{du_1}{dx} + \ldots$$

$$= (p^2 - u_0^2) - 2\left(\frac{\hbar}{i}\right)u_0u_1 - \left(\frac{\hbar}{i}\right)^2[u_1^2 + 2u_0u_2] + \ldots \tag{8.14}$$

---

4. Note that $p(x)$ defined by Eq. (8.11) is in agreement with the usual definition of the momentum of a particle.

Equating coefficients of like powers of $(\hbar/i)$ on either side of Eq. (8.14), we get,

$$u_0 = \pm p, \tag{8.15a}$$

$$u_1 = -\frac{1}{2u_0}\left(\frac{du_0}{dx}\right) = -\frac{1}{2p}\left(\frac{dp}{dx}\right). \tag{8.15b}$$

Corresponding to the two values of u_0, we have two values of u:

$$u_+ = p - \frac{\hbar}{i}\frac{1}{2p}\left(\frac{dp}{dx}\right) = p - \left(\frac{\hbar}{i}\right)\frac{d}{dx}(ln\sqrt{p}), \tag{8.16a}$$

$$u_- = -p - \frac{\hbar}{i}\frac{1}{2p}\left(\frac{dp}{dx}\right) = -p - \left(\frac{\hbar}{i}\right)\frac{d}{dx}(ln\sqrt{p}). \tag{8.16b}$$

Substituting in Eq. (8.9^2), we get,

$$\Phi_+(x) = \frac{1}{\sqrt{p}} \exp\left(\frac{i}{\hbar}\int^x p dx'\right) \tag{8.17a}$$

$$\Phi_-(x) = \frac{1}{\sqrt{p}} \exp\left(-\frac{i}{\hbar}\int^x p dx'\right) \tag{8.17b}$$

These represent the two independent solutions of the second order differential equation (8.3^2). The general solution, which is the WKB *wave function*, is given by a linear combination of Φ_+ and Φ_-:

$$\Phi_{WKB}(x) = \frac{A}{\sqrt{p}} \exp\left[\frac{i}{\hbar}\int^x p(x')dx'\right] + \frac{B}{\sqrt{p}} \exp\left[-\frac{i}{\hbar}\int^x p(x')dx'\right]. \tag{8.18}$$

The lower limit for the integral in (8.17) and (8.18) would be a classical turning point, as will be seen shortly.

Criterion for the Validity of the Approximation

The approximation leading to (8.18) is valid when condition (8.6b) is satisfied. In the present case, this condition reads,

$$|p^2| \gg \hbar\left|\frac{dp}{dx}\right|,$$

or

$$\frac{\lambda\,|(dp/dx)|}{|p|} \ll 1, \tag{8.19}$$

where, $\lambda = (\hbar/p)$, is the de-Bröglie wavelength of the particle. Thus, the condition for the applicability of the WKB approximation is that the fractional change of momentum over a de-Bröglie wavelength of the particle be small. This criterion could be compared with the criterion for the validity of ray (geometrical) optics: the variation of the index of refraction should not be appreciable over the distance of a wavelength.

Connection Formulae

Condition (8.19) is, obviously, not satisfied at a *classical turning point*, where $p(x) = 0$, but $\left(\frac{dp}{dx}\right) \neq 0$. The usefulness of the WKB method happens to be in the case of those problems that involve such turning points. That is, these are problems where we have two regions where the WKB approximation is valid but which are separated by a classical turning point (as in Fig. 8.1). The method can be applied to the solution of such problems only if we find a way to extend the WKB solution from one region to the other through the turning point. Such a procedure in effect, amounts to obtaining a *connection* between the WKB wavefunctions in the two regions which (the wavefunctions) can be written down independently of each other using formula (8.18). The procedure consists in solving the Schrödinger equation (8.3^2) exactly near the turning point and extrapolating the solution for regions far away from the turning point. The extrapolated (or, *asymptotic*) solutions will resemble the WKB solutions. Since the relationship between the extrapolated solutions in the two regions are known, we get the relationship between the two WKB solutions assuming this to be the same as that between the extrapolated solutions.

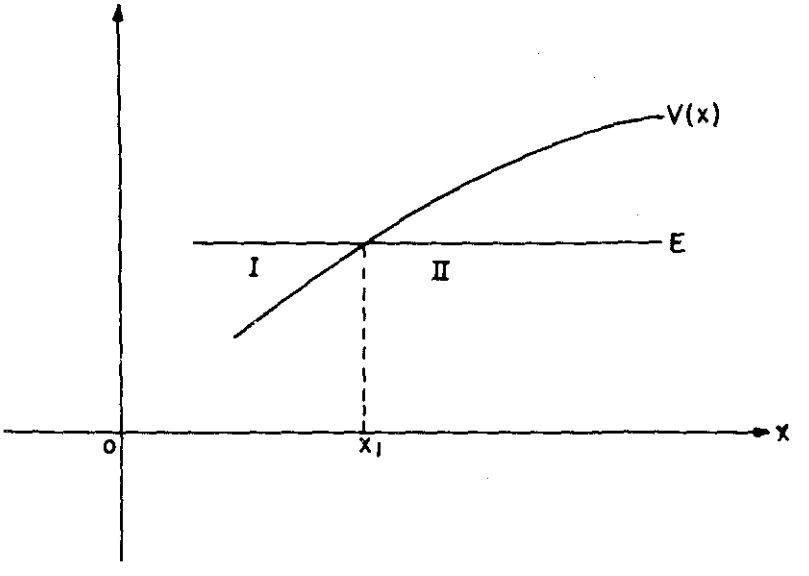


Fig. 8.1. Classical (I) and non-classical (II) regions separated by a classical turning point (x_1).

Now, a turning point separates a classical region, where $E > V(x)$, from a non-classical region where $E < V(x)$. The equations relating the WKB solutions

in the two regions, therefore, give the connection between the WKB wavefunctions in the classical and the non-classical regions. These equations are, for this reason, given the name *connection formulae.* A derivation of the formulae follows:

Consider a particle of energy E encountering a potential $V(x)$ as shown in Fig. 8.1. Classically, the particle would be turned back at $x = x_1$ (hence the name, *turning point*) where the kinetic energy $(E - V(x))$ becomes zero. But, in quantum mechanics, since the force $-(dV/dx)$ is finite at $x = x_1$, $V(x)$ represents a 'translucent' wall rather than an 'opaque' one (An opaque wall is represented by a potential which rises to infinity at $x = x_1$). This means that, at x_1, some particles will *leak* into region II (the non classical region), eventhough the majority might be turned back. This phenomenon is known as the *penetration of a potential barrier*, or, *tunnelling*.[5] In terms of an individual particle, we can say that there is a certain *probability* that it is found in region II if it was originally in region I. This probability can be estimated with the help of the WKB approximation.

In region I, $E > V(x)$, so $p(x)$ is real. The WKB wavefuction is, therefore, oscillatory. We have, from Eq. (8.18),

$$\Phi_I(x) = \frac{A_1}{\sqrt{p}} e^{+(i/\hbar)\int_{x_1}^{x} p\,dx'} + \frac{B_1}{\sqrt{p}} e^{-(i/\hbar)\int_{x_1}^{x} p\,dx'}$$

$$= \frac{A}{\sqrt{p}} \sin\left[-(1/\hbar)\int_{x_1}^{x} p\,dx' + \pi/4\right]$$

$$+ \frac{B}{\sqrt{p}} \cos\left[-(1/\hbar)\int_{x_1}^{x} p\,dx' + \pi/4\right], \qquad (8.20a)$$

where,

$$A = (A_1 - iB_1)e^{i\pi/4}, \qquad (8.21a)$$

$$B = -(A_1 + iB_1)e^{-i\pi/4}. \qquad (8.21b)$$

The reason for writing Φ_I in the peculiar form (8.20a) would be clear later.

In region II, $E < V(x)$ and so $p(x)$ is imaginary:

$$p(x) = i\,|\,p(x)\,|\,. \qquad (8.22)$$

Substituting in Eq. (8.18), we have, for the WKB wavefunction in region II, the expression,

$$\Phi_{II}(x) = \frac{A_2}{\sqrt{|\,p\,|}} e^{-(1/\hbar)\int_{x_1}^{x} |\,p\,|\,dx'} + \frac{B_2}{\sqrt{|\,p\,|}} e^{(1/\hbar)\int_{x_1}^{x} |\,p\,|\,dx'} \qquad (8.20b)$$

Thus, the WKB wavefunction of the particle is oscillatory in the classical region and is exponential in the non-classical region.

---

5. This name is inspired by the analogy of the crossing of a mountain by constructing a tunnel *through* it rather than by climbing *over* it.

Now, Φ_I and Φ_{II} are approximations to the same wavefunction. But we can identify them as such only if we know the relationship of the coefficients A_1 and B_1 (or A and B) in (8.20a) to the coefficients A_2 and B_2 in (8.20b). The connection formulae provide the required relationship. As we have already stated, the connection formulae are obtained by solving Eq. (8.3^2) exactly near the turning point and then finding the asymptotic forms of the solution far away on either side of the turning point.

We assume that $V(x)$ is *linear* in the neighbourhood of the turning point. That is,

$$V(x) \underset{x\sim x_1}{\approx} V(x_1) + (x - x_1)\left(\frac{dV}{dx}\right)_{x=x_1}$$

$$\approx E + C(x - x_1), \tag{8.23}$$

where,

$$C = \left(\frac{dV}{dx}\right)_{x=x_1} > 0 \tag{8.24}$$

Then,

$$p^2 = 2\mu(E - V) \underset{x\sim x_1}{\approx} 2\mu C(x - x_1). \tag{8.25}$$

Substituting for p^2 from (8.25), Eq. (8.3^2) reduces to,

$$\frac{d^2\psi}{d\xi^2} - \xi\psi = 0, \tag{8.26}$$

where,

$$\xi = (2\mu C/\hbar^2)^{1/3}(x - x_1), \tag{8.27}$$

and

$$\psi(\xi) = \Phi(x).$$

Regions I and II and the turning point correspond, respectively, to $\xi < 0$, $\xi > 0$ and $\xi = 0$.

The solutions of Eq. (8.26) are known as the *Airy Functions*[6], and are given by,

$$Ai(\xi) = \frac{1}{\pi}\int_0^\infty \cos\,(s^3/3 + s\xi)\,ds, \tag{8.28}$$

and

$$Bi(\xi) = \frac{1}{\pi}\int_0^\infty \{\exp(-s\xi - s^3/3) + \sin\,(s^3/3 + s\xi)\}\,ds. \tag{8.29}$$

---

6. Jeffreys, H. and B.S., *Methods of Mathematical Physics* (Cambridge University Press, Cambridge 1956), III Edition, Section 17.07.

We are interested only in the asymptotic forms of Ai and Bi. These are given by,

$$Ai(\xi) \underset{\xi<<0}{\sim} (-\pi^2\xi)^{-1/4}\sin[(2/3)(-\xi)^{3/2}+\pi/4], \tag{8.28^1}$$

$$\underset{\xi>>0}{\sim} \frac{1}{2}(\pi^2\xi)^{-1/4}\exp[-(2/3)\xi^{3/2}]. \tag{8.28^2}$$

$$Bi(\xi) \underset{\xi<<0}{\sim} (-\pi^2\xi)^{-1/4}\cos[(2/3)(-\xi)^{3/2}+\pi/4], \tag{8.29^1}$$

$$\underset{\xi>>0}{\sim} (\pi^2\xi)^{-1/4}\exp[(2/3)\xi^{3/2}]. \tag{8.29^2}$$

Now, for $\xi < 0$ (that is, for $x < x_1$),

$$\begin{aligned}(2/3)(-\xi)^{3/2} &= \int_0^{+\xi}\sqrt{-\xi'}\,d(-\xi') \\ &= -\int_0^{\xi}\sqrt{-\xi'}\,d\xi' \\ &= -\int_{x_1}^{x}(2\mu C/\hbar^2)^{1/2}(x_1-x')^{1/2}dx' \\ &= \frac{-1}{\hbar}\int_{x_1}^{x}p(x')dx',\end{aligned} \tag{8.30a}$$

where, Eqs. (8.27) and (8.25) are used in the last two lines.

Also,

$$\begin{aligned}(-\xi)^{1/4} &= (2\mu C)^{1/12}\hbar^{-1/6}(x_1-x)^{1/4} \\ &= (2\mu C\hbar)^{-1/6}[2\mu C(x_1-x)]^{1/4} \\ &= (2\mu C\hbar)^{-1/6}\sqrt{p(x)}.\end{aligned} \tag{8.31a}$$

Similarly, for $\xi > 0$ ($x > x_1$), we have,

$$(2/3)\xi^{3/2} = \frac{1}{\hbar}\int_{x_1}^{x}|\,p(x')\,|\,dx', \tag{8.30b}$$

and

$$(\xi)^{1/4} = (2\mu C\hbar)^{-1/6}\sqrt{|\,p(x)\,|}. \tag{8.31b}$$

Substituting from Eqs. (8.30a-31b) in Eqs. (8.28^1-29^2), we get,

$$Ai(\xi) \underset{\xi<<0}{=} \Phi_1^{osc}(x) \underset{x<<x_1}{\sim} \frac{\alpha}{\sqrt{p}}\sin\left[+\frac{1}{\hbar}\int_x^{x_1}p(x')dx'+\frac{\pi}{4}\right], \tag{8.32^1}$$

$$Ai(\xi) \underset{\xi>>0}{=} \Phi_1^{exp}(x) \underset{x>>x_1}{\sim} \frac{\alpha}{2\sqrt{|\,p\,|}}\exp\left(-\frac{1}{\hbar}\int_{x_1}^{x}|\,p\,|\,dx'\right), \tag{8.32^2}$$

$$Bi(\xi) \underset{\xi \ll 0}{=} \Phi_2^{osc}(x) \underset{x \ll x_1}{\sim} \frac{\alpha}{\sqrt{p}} \cos\left[\frac{1}{\hbar}\int_x^{x_1} p(x')dx' + \frac{\pi}{4}\right], \tag{8.33^1}$$

$$Bi(\xi) \underset{\xi \gg 0}{=} \Phi_2^{exp}(x) \underset{x \gg x_1}{\sim} \frac{\alpha}{\sqrt{|p|}} \exp\left(\frac{1}{\hbar}\int_{x_1}^{x} |p|\,dx'\right), \tag{8.33^2}$$

Here,

$$\alpha = [2\mu C\hbar/\pi^3]^{1/6}. \tag{8.34}$$

Since Φ_k^{exp} and Φ_k^{osc} are approximations to the same wave-function, the former is the continuation into the non-classical region of the latter in the classical region. Thus, the connection between the approximate forms of the wavefunctions in the classical and the non-classical regions are given by the formulae:

Classical region (oscillatory)		*Non-classical region* (exponential)
$\alpha p^{-1/2}\sin\left[\frac{1}{\hbar}\int_x^{x_1} pdx' + \frac{\pi}{4}\right]$	$\leftrightarrow$	$\frac{\alpha}{2}\lvert p\rvert^{-1/2}\exp\left\{\frac{1}{\hbar}\int_x^{x_1} \lvert p\rvert\,dx'\right\}$, (8.35a)
$\beta p^{-1/2}\cos\left[\frac{1}{\hbar}\int_x^{x_1} pdx' + \frac{\pi}{4}\right]$	$\leftrightarrow$	$\beta\lvert p\rvert^{-1/2}\exp\left\{\frac{1}{\hbar}\int_{x_1}^{x} \lvert p\rvert\,dx'\right\}$. (8.35b)

We note that a wavefunction that is represented by the sine function in the classical region becomes a *decreasing* exponential in the non-classical region whereas an *increasing* exponential in the non-classical region corresponds to the cosine function in the classical region. Note also that the constant multiplying the increasing exponential is the same as that multiplying the cosine function whereas the constant multiplying the decreasing exponential is half that associated with the sine function.

The wavefunction of the physical system would be a general solution of Eq. (8.26) and, thus, a linear combination of Φ_1 and Φ_2. Thus,

$$\Phi^{osc}(x) \underset{x \ll x_1}{\sim} \frac{\alpha}{\sqrt{p}}\left[\sin\left\{\frac{1}{\hbar}\int_x^{x_1} p(x')dx' + \frac{\pi}{4}\right\} + \cos\left\{\frac{1}{\hbar}\int_x^{x_1} p(x')dx' + \frac{\pi}{4}\right\}\right], \tag{8.36a}$$

$$\Phi^{exp}(x) \underset{x \gg x_1}{\sim} \frac{(\alpha/2)}{\sqrt{|p|}} \exp\left(-\frac{1}{\hbar}\int_{x_1}^{x} |p(x')|\,dx'\right) + \frac{\alpha}{\sqrt{|p|}} \exp\left(\frac{1}{\hbar}\int_{x_1}^{x} |p(x')|\,dx'\right). \tag{8.36b}$$

Comparing Eqs. (8.36a, b) with (8.20a, b), we see that Φ^{osc} is the WKB wavefunction Φ_I in the classical region and Φ^{exp} is the WKB wavefunction Φ_{II} in the non-classical region. We further see that

$$A_2 = (\alpha/2) = A/2,$$

$$B_2 = \alpha = B = \left\{ \frac{2\mu\hbar}{\pi^3} \left(\frac{dV}{dx} \right)_{x=x_1} \right\}^{1/6} \tag{8.37}$$

Thus, Eqs. (8.35a, b) give the connection between the WKB wavefunctions in the classical (oscillatory) and the non-classical (exponential) regions. These are, therefore, the connection formulae of the WKB approximation.

Since the approximations (8.32[1]-8.33[2]) are valid only for regions far away from the turning points, the above method cannot be applied when there are two turning points close to each other as in Fig. 8.2. In fact, substituting from (8.25) in (8.19), we get,

$$| x - x_1 | \gg (\lambda/2\pi) \tag{8.38}$$

as the condition for the validity of the WKB approximation at x. Thus, for the applicability of the WKB method, it is necessary that the separation between turning points be at least several de-Bröglie wavelengths.

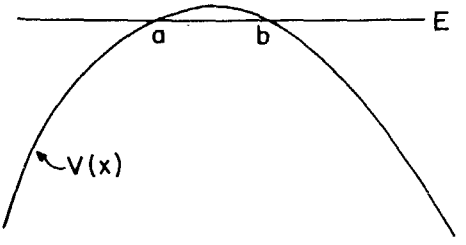


Fig. 8.2. The two classical turning points (a and b in the figure) are too near to each other for the applicability of the WKB approximation.

APPLICATIONS

A. Bound State

The WKB method can be usefully applied for the determination of the energy levels of a one-dimensional bound system. The potential for such a system is represented by the curve $V(x)$ in Fig. 8.3. There are three regions separated by the two turning points x_1 and x_2, as shown in the figure[7]. Region II is the classical region where the wavefunction is oscillatory, whereas regions I and III are the non-classical (exponential) ones. Since the system is bound, the wavefunction should go to zero as $x \to \pm\infty$. This means that, in regions I and III, the WKB wavefunctions are *decreasing* exponentials.
Thus, by eq. (8.18),

$$\Phi_I(x) \underset{x<a_1}{\approx} \frac{A_1}{\sqrt{| p |}} \exp\left\{ -\frac{1}{\hbar} \int_x^{x_1} | p(x') | \, dx' \right\}, \tag{8.39}$$

7. The regions I, II and III, where the WKB approximation is valid are actually defined by $x \leq a_1, b_1 \leq x \leq b_2$ and $x \geq a_2$, respectively [see, inequality (8.38)].

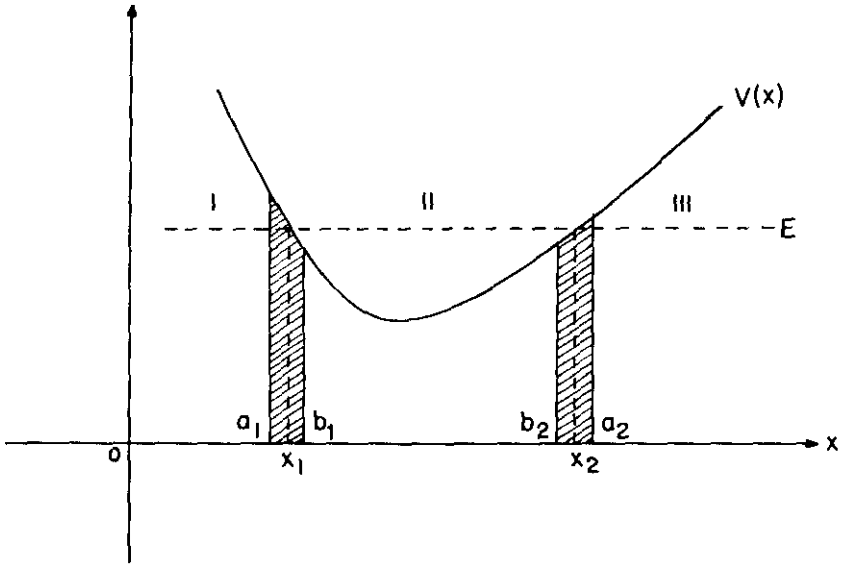


Fig. 8.3. Potential well. The shaded portions denote regions where the WKB approximation is not valid.

$$\Phi_{III}(x) \underset{x>a_2}{\approx} \frac{A_3}{\sqrt{|p|}} \exp\left\{-\frac{1}{\hbar}\int_{x_2}^{x} |p(x')|\,dx'\right\}. \tag{8.40}$$

According to the connection formulae (8.35a, b) the wavefunction in region II is given by

$$\Phi_{II}(x) \approx \frac{2A_1}{\sqrt{p}} \sin\left\{\frac{1}{\hbar}\int_{x_1}^{x} p(x')dx' + \pi/4\right\} \tag{8.41a}$$

$$\approx \frac{2A_3}{\sqrt{p}} \sin\left\{\frac{1}{\hbar}\int_{x}^{x_2} p(x')dx' + \pi/4\right\}. \tag{8.41b}$$

But
$$\int_{x_1}^{x} p\,dx' = \int_{x_1}^{x_2} p\,dx' - \int_{x}^{x_2} p\,dx',$$

so that,

$$\sin\left\{\frac{1}{\hbar}\int_{x_1}^{x} p\,dx' + \pi/4\right\} = \sin\left\{\left(\frac{1}{\hbar}\int_{x_1}^{x_2} p\,dx' + \pi/2\right) - \left[\frac{1}{\hbar}\int_{x}^{x_2} p\,dx' + \pi/4\right]\right\}. \tag{8.42}$$

Substituting (8.42) in (8.41a) and equating the result to (8.41b), we get,

$$A_1 \sin\left\{\left(\frac{1}{\hbar}\int_{x_1}^{x_2} p\,dx' + \pi/2\right) - \left[\frac{1}{\hbar}\int_{x}^{x_2} p\,dx' + \pi/4\right]\right\}$$

$$= A_3 \sin\left\{\frac{1}{\hbar}\int_x^{x_2} p\, dx' + \pi/4\right\}.$$

Comparing this with the identity,

$$\sin(n'\pi - \theta) = (-1)^{n'-1}\sin\theta,\ n' = 1, 2, 3, \ldots \tag{8.43}$$

we should have,[8]

$$\left.\begin{aligned} \frac{1}{\hbar}\int_{x_1}^{x_2} p(x)dx + \pi/2 &= (n+1)\pi, \\ A_3/A_1 &= (-1)^n, \end{aligned}\right\} n = 0, 1, 2, \ldots \tag{8.44}$$

Now,

$$2\int_{x_1}^{x_2} p\, dx = \int_{x_1}^{x_2} p\, dx - \int_{x_2}^{x_1} p\, dx = \oint p\, dx, \tag{8.45}$$

where $\oint$ represents integration over a complete period (or to-and-fro motion) of the particle. Substituting from (8.45) in (8.44), we get,

$$\oint p\, dx = \left(n + \frac{1}{2}\right)h,\ (n = 0, 1, 2, \ldots) \tag{8.46}$$

When the potential is known as a function of x, the integral in Eq. (8.46) can be performed, yielding E in terms of n and the parameters of the potential. n, thus designates the different energy levels of the system.

Eq. (8.46) corresponds to the *Bohr-Sommerfeld quantization rule* of the Old Quantum Theory (of the pre-quantum mechanics period). The term $\frac{1}{2}h$, which is absent in the Bohr-Sommerfeld formula, brings formula (8.46) in better agreement with the exact result. In fact, in the case of the linear harmonic oscillator, Eq. (8.46) is in agreement with the exact result [see Eq. (4.51) and Problem 8.1].

The approximate wavefunction of the system is given by Eq. (8.41a). The constant A_1 can be determined from the requirement of normalization. According to Eq. (8.44), the phase of the sine function in (8.41a) varies from $\pi/4$ to $(n + 3/4)\pi$ as x varies from x_1 to x_2. Thus, n is the number of zeroes (that is, *nodes*) of $\Phi_n(x) \equiv \Phi_{II}(x)$ between x_1 and x_2. But the WKB approximation is valid only at distances that are several de Bröglie wavelengths (a wavelength being twice the distance between nodes) removed from the turning points [Eq. (8.38)]. This means that Eq. (8.41a) is a good approximation to the wavefunction of the system only for large values of the quantum number n. In that case, the sine function oscillates rapidly in the interval $x_1 < x < x_2$ so that the square of the sine function can be approximated to its average value $\frac{1}{2}$.

---

8. We have $(n+1)\pi$ rather than $n\pi$ on the R.H.S. of Eq. (8.44), because $\int_{x_1}^{x_2} p\, dx \geq 0$.

Then,

$$1 = \int_{-\infty}^{+\infty} |\Phi_n(x)|^2 dx \approx \int_{x_1}^{x_2} |\Phi_n(x)|^2 dx.$$

$$\approx 4|A_1|^2 \int_{x_1}^{x_2} \frac{dx}{2p(x)} = 4|A_1|^2 (\tau_n/4\mu)$$

$$= |2A_1|^2 (\pi/2\mu\omega_n), \tag{8.47}$$

where, $\tau_n = (2\pi/\omega_n) = 2\mu \int_{x_1}^{x_2} \left(\frac{dx}{p}\right)$, is the period (the time required for the particle to move from x_1 to x_2 and x_2 to x_1) of the nth mode, ω_n being the angular frequency. Substituting (8.47) in (8.41a), we have, for the normalized wavefunction, the expression,

$$\Phi_n(x) \approx \left(\frac{2\mu\omega_n}{\pi p}\right)^{1/2} \sin\left\{\frac{1}{\hbar}\int_{x_1}^{x} p(x')dx' + \pi/4\right\} \tag{8.48}$$

Problem 8.1: Show, from Eq. (8.44), that the energy levels of a linear harmonic oscillator $\left[V(x) = \frac{1}{2}\mu\omega^2 x^2\right]$, for large values of the quantum number n, are given by

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega. \tag{8.49}$$

Eq. (4.51) shows that the WKB result (8.49) is, in the case of the linear harmonic oscillator, exact, valid for all values of n.

B. Penetration of a Potential Barrier

Another situation where the WKB approximation can be used with advantage, is in the calculation of the transmission coefficient of a potential barrier of the type shown in Fig. 8.4.

Particles of energy E less than the height of the potential barrier, are incident from the left of the barrier. At the classical turning point x_1 (defined by $V(x_1) = E$), some of the particles will pass on to the classically-forbidden region defined by $V(x) > E$. Of these, some will be reflected back at the other turning point[9] x_2 (where $V(x_2) = E$), but the others will escape to the classical region to the right of the potential barrier. Thus, there is a possibility that a certain fraction (which would, naturally, depend on the parameters such as height (measured from E), width and shape of the barrier) of the total number of incident particles would be *transmitted* by the potential. The ratio of the flux of transmitted particles to that of the incident particles, is called the *transmission coefficient* of the potential barrier. The WKB approximation enables us to obtain an expression for the transmission coefficient in terms of the above mentioned parameters of the potential barrier.

---

9 x_2 is the point at which particles incident on the barrier *from the right* would be turned back, according to the laws of classical mechanics.

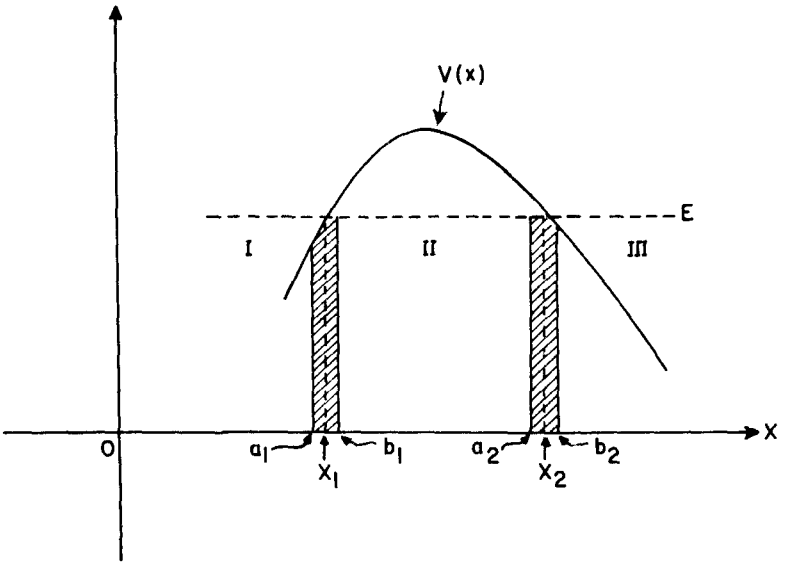


Fig. 8.4. Potential barrier.

We denote by I, II and III the regions of validity of the WKB approximation. These are defined, respectively, by (see Fig. 8.4), $x \leq a_1, b_1 \leq x \leq a_2$ and $x \geq b_2$. The WKB wavefunction in region I is given, according to Eq. (8.18), by

$$\Phi_I(x) = \frac{A_1}{\sqrt{p}} \exp\left[\frac{i}{\hbar}\int_{x_1}^{x} p(x')dx'\right] + \frac{B_1}{\sqrt{p}} \exp\left[-\frac{i}{\hbar}\int_{x_1}^{x} p(x')dx'\right]. \tag{8.50}$$

Here, the first term on the R.H.S. represents the incident particles and the second term particles reflected at the turning point x_1. This could be easily seen by shifting the origin to x_1 and taking the special case where p is constant.

In region III, there is only transmitted wave, so that

$$\Phi_{III} = \frac{A_3}{\sqrt{p}} \exp\left[(i/\hbar)\int_{x_2}^{x} p(x')dx'\right]. \tag{8.51}$$

From Eqs. (8.50) and (8.51), we have, for the transmission coefficient T, the expression,

$$T = \frac{|\mathbf{J}_{tr}|}{|\mathbf{J}_{inc}|} = \frac{|\rho_{tr}\mathbf{v}_{tr}|}{|\rho_{inc}\mathbf{v}_{inc}|} = \frac{|\psi_{tr}\sqrt{p_{tr}}|^2}{|\psi_{inc}\sqrt{p_{inc}}|^2}$$

$$\approx \frac{|\sqrt{p}\,\Phi_{III}|^2}{|\sqrt{p}(\Phi_I)_{inc}|^2} = \frac{|A_3|^2}{|A_1|^2}. \tag{8.52[1]}$$

Here, the symbols $\mathbf{J}$, ρ and $\mathbf{v}$ stand, respectively, for the flux, density and velocity of the particles, while ψ represents the (exact) wave function. Also, the fact that the momenta of the incident and the transmitted particles are the same, has been made use of.

Thus, if A_1 is determined in terms of A_3, T can be obtained. The relationship between A_1 and A_3 can, indeed, be found using the connection formulae (8.35a, b), as indicated below:

Rewriting Eq. (8.51) as,

$$\Phi_{III}(x) = \frac{A}{\sqrt{p}}\left\{\cos\left[\frac{1}{\hbar}\int_{x_2}^{x} p(x')dx' + \frac{\pi}{4}\right] + i\sin\left[\frac{1}{\hbar}\int_{x_2}^{x} p(x')dx' + \frac{\pi}{4}\right]\right\},$$

where,

$$A = A_3 e^{-i\pi/4} = \left(\frac{1-i}{\sqrt{2}}\right)A_3, \tag{8.53}$$

we get, for the WKB wave function in region II,

$$\Phi_{II}(x) = \frac{A}{\sqrt{|p|}}\left\{\left(\exp\left[(1/\hbar)\int_{x}^{x_2}|p|\,dx'\right]\right) + (i/2)\left(\exp\left[-(1/\hbar)\int_{x}^{x_2}|p|\,dx'\right]\right)\right\}. \tag{8.54a}$$

Writing,

$$\int_{x}^{x_2}|p|\,dx' = \int_{x_1}^{x_2}|p|\,dx' - \int_{x_1}^{x}|P|\,dx'$$

and defining

$$\Theta = \exp\left[(1/\hbar)\int_{x_1}^{x_2}|p|\,dx'\right] \tag{8.55}$$

this becomes,

$$\Phi_{II}(x) = \frac{A}{\sqrt{|p|}}\left\{\Theta\exp\left[-(1/\hbar)\int_{x_1}^{x}|p|\,dx'\right] + (i/2\Theta)\exp\left[(1/\hbar)\int_{x_1}^{x}|p|\,dx'\right]\right\}. \tag{8.54b}$$

Applying the connection formulae again, we get,

$$\Phi_1 = \frac{A}{\sqrt{p}}\left\{2\Theta\sin\left[\frac{1}{\hbar}\int_{x}^{x_1} p\,dx' + \pi/4\right] + (i/2\Theta)\cos\left[\frac{1}{\hbar}\int_{x}^{x_1} p\,dx' + \pi/4\right]\right\}$$

$$= \frac{A_3}{\sqrt{p}}\left\{\left(\Theta + \frac{1}{4\Theta}\right)\exp\left[(i/\hbar)\int_{x_1}^{x} p\,dx'\right] - i\left(\Theta - \frac{1}{4\Theta}\right)\exp\left(-\frac{i}{\hbar}\int_{x_1}^{x} p\,dx'\right)\right\}, \tag{8.56}$$

where, relationship (8.53) has been used.

Comparing Eqs. (8.50) and (8.56), we get,

$$A_1 = \left(\Theta + \frac{1}{4\Theta}\right) A_3 , \tag{8.57a}$$

$$B_1 = -i\left(\Theta - \frac{1}{4\Theta}\right) A_3 . \tag{8.57b}$$

Thus

$$T = |A_3/A_1|^2 = \left(\Theta + \frac{1}{4\Theta}\right)^{-2} \approx (1/\Theta)^2 = \exp\left\{\frac{-2}{\hbar}\int_{x_1}^{x_2} |p(x)|\, dx\right\}. \tag{8.52^2}$$

Θ defined by Eq. (8.55) is a measure of both the height, $|p(x)| = \{2\mu(V(x)-E)\}^{1/2}$, and the width $(x_2 - x_1)$ of the barrier. Since the WKB approximation is valid only when $(x_2 - x_1)$ is several times the wave length $\lambda = \hbar/|p|, \Theta \gg 1$. Hence the approximation $\left(\Theta + \frac{1}{4\Theta}\right)^2 \approx \Theta^2$, in Eq. (8.52^2).

We see from Eq. (8.52^2), that increasing either the width or the height of the barrier, decreases the probability for the penetration of the barrier. This is as it should be. For, $x_1 < x < x_2$ represents a region where the beam of particles lose intensity continually, this loss being greater, the greater the height of the barrier. As a result, the chance for a particle to reach the barrier boundary x_2 decreases with increasing distance of x_2 from x_1 as well as with increasing height of the barrier.

Problem 8.2: Calculate the transmission coefficient of the potential barrier given by

$$V(x) = V_0(1 - x^2/a^2), \text{ for } |x| \le a$$

$$= 0 \qquad\qquad , \text{for } |x| > a$$

Potential with a Vertical Wall

In the case of a potential with a vertical wall, as shown in Fig. 8.5, the linear approximation (8.23) would not hold good at the turning point x_1. As a result, the connection formulae (8.35) are not applicable, without necessary modification, at this turning point. The necessary modification can be found out by solving the Schrödinger equation exactly in the region I. The WKB approximations are assumed to hold good for regions II and III.

As an example, consider the potential shown in Fig. 8.5(a). It is given by,

$$V(x) = Cx\ ,\ x > 0, \tag{8.58}$$

$$= +\infty,\ \text{at } x = 0.$$

Since $V = +\infty$ represents a perfectly opaque wall, the wave function vanishes in region I, including on the wall. The continuity of the wave function at $x = x_1$ requires, then, that $\Phi_{II}(x_1)$ also be zero. Since the WKB wavefunction in region III is given by Eq. (8.32^2) with x_1 replaced by x_2, we see from (8.32^1) that the WKB wavefunction in region II should be given by

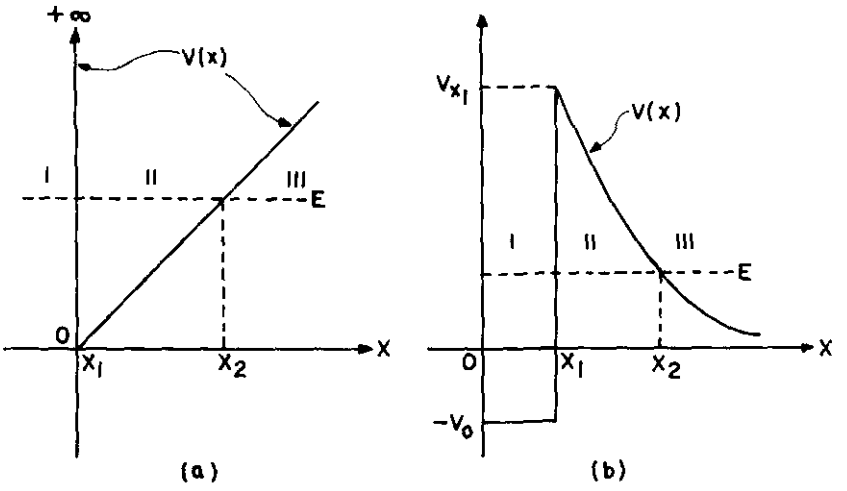


Fig. 8.5. Potential with vertical walls. The linear approximation, Eq. (8.23), is not valid at the turning point x_1.

$$\Phi_{II}(x) \approx \frac{A}{\sqrt{p}} \sin\left\{-\frac{1}{\hbar}\int_0^x p\, dx'\right\}, \tag{8.59}$$

with $p(x) = \{2\mu(E - Cx)\}^{1/2}$.

This modifies the connection formula (8.35a) at the turning point x_2 to:

$$\frac{2}{\sqrt{p}} \sin\left\{-\frac{1}{\hbar}\int_{x_2}^x p\, dx'\right\} \leftrightarrow \frac{1}{\sqrt{|p|}} \exp\left(-\frac{1}{\hbar}\int_{x_2}^x |p|\, dx'\right). \tag{8.35a'}$$

This could be regarded as the connection formula whenever the wavefunction at the other turning point vanishes.

As another example, let us consider the potential shown in Fig. 8.5(b), which is given by[10],

$$V(x) = \beta x^{-1},\ x \geq x_1,$$
$$= -V_0,\ 0 \leq x \leq x_1. \tag{8.60}$$

The Schrödinger equation in region I is given by

$$\frac{d^2\Phi}{dx^2} + k^2\Phi = 0, \tag{8.61}$$

where,

$$k^2 = 2\mu(V_0 + E)/\hbar^2. \tag{8.62}$$

Eq. (8.61) has the exact solution,

10. (8.60), roughly, corresponds to the potential barrier faced by an Alpha particle inside the nucleus in an Alpha particle decay.

$$\Phi_1(x) = B \sin\{k(x - x_1) + \delta\}$$

$$= \frac{A_1}{\sqrt{p}} \exp\left[i\{k(x - x_1) + \delta\}\right] + \frac{B_1}{\sqrt{p}} \exp\left[-i\{k(x - x_1) + \delta\}\right], \tag{8.63}$$

where B and δ are constants, and

$$A_1 = -(i/2)B\sqrt{p(x_1)}\,;\; B_1 = (i/2)B\sqrt{p(x_1)}, \tag{8.64}$$

The WKB wavefunctions in regions II and III are given by Eqs. (8.54b) and (8.51), respectively. From the continuity of the wavefunction and its logarithmic derivative $\left(\frac{d\Phi}{dx}\right)/\Phi$ at $x = x_1$, we get (since $\Theta \gg 1$),

$$B \sin\delta = \frac{A}{\sqrt{|\,p(x_1)\,|}}\left\{\Theta + \frac{i}{2\Theta}\right\} \approx \frac{\Theta A}{\sqrt{|\,p(x_1)\,|}}, \tag{8.65a}$$

$$k \cot\delta = \frac{-|\,p(x_1)\,|}{\hbar}\left(\frac{\Theta - i/2\Theta}{\Theta + i/2\Theta}\right) \approx \frac{-|\,p(x_1)\,|}{\hbar}, \tag{8.65b}$$

with $|\,p(x_1)\,| = \{2\mu(V_{x_1} - E)\}^{1/2}$, (8.66)

where V_{x_1} is shown in Fig. 8.5 (b).

From Eqs. (8.53) and (8.65a), we have,

$$A_3 = \frac{1+i}{\sqrt{2}}A = \left(\frac{1+i}{\sqrt{2}}\right)\frac{B\{|\,p(x_1)\,|\}^{1/2}\sin\delta}{\Theta}, \tag{8.67}$$

and from Eqs. (8.52^1), (8.64), (8.67) and (8.65b), we get,

$$T = \left|\frac{A_3}{A_1}\right|^2 = (4\sin^2\delta)\Theta^{-2}$$

$$= \left(\frac{4}{1 + \cot^2\delta}\right)\Theta^{-2}$$

$$= \left\{\frac{4(V_0 + E)}{V_0 + V_{x_1}}\right\}\Theta^{-2}, \tag{8.68}$$

where the relationship $\frac{|\,p(x_1)\,|}{p(x_1)} = -1$, as well as Eq. (8.62), have also been used.

Comparing Eq. (8.68) with (8.52^2), we see that, in most cases of interest, the transmission coefficient is increased for a potential with a discontinuity at x_1 as compared with that for a potential which is linear at x_1. Formula (8.68) yields a value of $T = 4$ when $E = V_{x_1}$ in place of the exact value of $T \approx 1$.

Problem 8.3: Determine, in the WKB approximation, the energy levels of a particle moving in a uniform gravitational field when the motion is limited from below by a perfectly reflecting plane.

8.2 THE VARIATIONAL METHOD

8.2A Bound States (Ritz Method)

Often, the basic dynamical equations of physics can be derived from a variational principle. Thus, the Lagrange's equation of motion in classical mechanics follow from Hamilton's principle of least action[11] [Eq. (1.7)], whereas the eikonal equation in geometrical optics[12] derives from Fermat's principle of least time. We will see in chapter 11 (Section 11.2), that the classical field equations are derivable from a variational principle. It is, therefore, not surprising that the basic equation of quantum mechanics (the Schrödinger equation) also is equivalent to a variational equation.

Now, for a variational principle, we require a function which would be 'stationary' with respect to variations of its parameters.[13] In the case of the Lagrange's equations, this function is the action S defined by Eq. (1.6). For the case of the Schrödinger equation, the function turns out to be the expectation value $E(\psi)$ defined by [See Eq. (3.2a)],

$$E(\psi) = \frac{<\psi \mid \hat{H} \mid \psi>}{<\psi \mid \psi>}, \tag{8.69}$$

where $\hat{H}$ is the Hamiltonian of the system and ψ is the variational wave function (the *trial function*). It is easily shown (Problem 8.4) that the variation in $E(\psi)$ linear in $\delta\psi$ is indeed zero for an appropriate choice of ψ. Eq. (8.69) is referred to as the *Ritz* (or, *Rayleigh-Ritz*) *variational formula*, while the variational method based on it is known as the *Ritz method.*

Now, the stationary values of $E(\psi)$ are given by the variational equation,

---

11. See, for example, Landau, L.D. and Lifshitz, E.M., *Mechanics* (Pergamon Press, 1969), Section 2.
12. Born, M. and Wolf, E. *Principles of Optics*, II Edition (Pergamon Press, 1964), Appendix I. II.
13. If F is the function and α is one of the parameters, then the variation δF in F corresponding to the infinitesimal variation $\delta\alpha = (\alpha - \alpha_0)$ in α (keeping the other parameters constant) can be written as

$$\delta F = F(\alpha) - F(\alpha_0)$$

$$= \left(\frac{\partial F}{\partial \alpha}\right)_{\alpha_0} \delta\alpha + \left(\frac{\partial^2 F}{\partial \alpha^2}\right)_{\alpha_0} \frac{(\partial\alpha)^2}{2!} + \ldots \approx \left(\frac{\partial F}{\partial \alpha}\right)_{\alpha_0} \delta\alpha.$$

Thus, F is stationary at $\alpha = \alpha_0$ if

$$\left(\frac{\partial F}{\partial \alpha}\right)_{\alpha_0} = 0;$$

that is, if $F(\alpha_0)$ is either a maximum or a minimum.

$$\delta E(\psi) = 0. \tag{8.70}$$

It is to be shown, therefore, that Eq. (8.70) is equivalent to the time-independent Schrödinger equation. For this, let us write Eq. (8.69) in the form,

$$\langle \psi \mid \psi \rangle E = \langle \psi \mid \hat{H} \mid \psi \rangle.$$

Taking the variation of this equation corresponding to the variation $\delta\psi$ in ψ, we have,

$$\langle \psi \mid \psi \rangle \delta E + \{\langle \delta\psi \mid \psi \rangle + \langle \psi \mid \delta\psi \rangle\} E = \langle \delta\psi \mid \hat{H} \mid \psi \rangle + \langle \psi \mid \hat{H} \mid \delta\psi \rangle,$$

or,
$$\langle \psi \mid \psi \rangle \delta E = \langle \delta\psi \mid (\hat{H} - E) \mid \psi \rangle + \langle \delta\psi \mid (\hat{H} - E) \mid \psi^* \rangle. \tag{8.71}$$

Hence, when $\delta E = 0$, we have,

$$\text{Re}\,\{\langle \delta\psi \mid (\hat{H} - E) \mid \psi \rangle\} = 0, \tag{8.72a}$$

where, Re { } represents the real part. Similarly, replacing $\delta\psi$ by $\delta\psi' = i\delta\psi$, we get,

$$-i[\langle \delta\psi \mid (\hat{H} - E) \mid \psi \rangle - \langle \delta\psi \mid (\hat{H} - E) \mid \psi \rangle^*] = 0,$$

or,

$$\text{Im}\,[\langle \delta\psi \mid (\hat{H} - E) \mid \psi \rangle] = 0. \tag{8.72b}$$

From Eqs. (8.72a, b), we have[14],

$$\langle \delta\psi \mid (\hat{H} - E) \mid \psi = 0. \tag{8.72}$$

This relationship can be satisfied for arbitrary $\delta\psi$ only if

$$(\hat{H} - E) \mid \psi \rangle = 0,$$

or

$$\hat{H}\psi = E\psi, \tag{8.73}$$

which is the Schrödinger equation for stationary states [Eq. (4.18)].

Thus, the solutions of Eq. (8.70) are solutions of Eq. (8.73).

Problem 8.4: A trial function for a variational calculation is of the form $\psi = \phi_0 + \epsilon\, \phi_1$, where $|\epsilon| \ll 1$ and ϕ_0 and ϕ_1 are normalized. Obtain the conditions on ϕ_0 and ϕ_1 for $E(\psi)$ defined by Eq. (8.69) to be stationary to first order in ϵ .

Now, the solutions of Eq. (8.70) correspond to either maxima or minima of the function $E(\psi)$ (see Footnote 13). In fact, the solution corresponds to a minimum:

Let $\{\phi_k\}$ represent the complete orthonormal set of eigenfunctions of $\hat{H}$. Then ψ can be expanded in terms of $\{\phi_k\}$.

$$| \psi \rangle = \sum_j c_j \mid \phi_j \rangle, \tag{8.74}$$

$$E(\psi) = \frac{\sum_j | c_j |^2 E_j}{\sum_j | c_j |^2} \geq E_0. \tag{8.75}$$

---

14. Eq. (8.72) will follow from Eq. (8.71) if we regard the variation $| \delta\psi \rangle$ and $\langle \delta\psi |$ in the latter as linearly independent of each other. Such a viewpoint would be justified since the relationship between a ket vector and a bra vector is an antilinear one (see Section 2.3).

where E_j is the eigenvalue of $\hat{H}$ belonging to ϕ_j and E_0 is the ground state energy (the smallest E_j).

$E(\psi)$ is, obviously, independent of the normalization of ψ and, therefore, we can use a normalized trial function. The method would consist in evaluating $E(\psi)$ with a trial function and then varying ψ until $E(\psi)$ is a minimum. Let ψ_0 be the ψ which corresponds to this minimum. Then, according to (8.75), $E(\psi_0) \geq E_0$. Thus, the procedure yields an upper limit to the ground state energy of the system and an approximate ground state wavefunction ψ_0. The success (the accuracy) of the method depends on the correct choice (at least in form) of the trial function. Thus, the method is truly an approximation method since the trial function, whose choice is usually based on circumstantial factors, is rarely likely to be an exact wavefunction of the system.

In practice, the trial function is defined in terms of a number of unknown parameters $\alpha, \beta, \ldots$:

$$\psi \equiv \psi(\mathbf{r}; \alpha, \beta, \ldots),$$

so that,

$$E(\psi) \equiv \langle \psi \mid \hat{H} \mid \psi \rangle = \int \psi^*(\mathbf{r}; \alpha, \beta \ldots) \hat{H} \psi(\mathbf{r}; \alpha, \beta, \ldots) d^3\mathbf{r}$$

$$= J(\alpha, \beta, \ldots), \text{ say .} \tag{8.76}$$

Then

$$\psi_0 \equiv \psi(\mathbf{r}; \alpha_0, \beta_0, \ldots), \tag{8.77a}$$

and

$$E(\psi_0) = J(\alpha_0, \beta_0, \ldots), \tag{8.77b}$$

where, $\alpha_0, \beta_0, \ldots$ are values of $\alpha, \beta, \ldots$ for which the integral (8.76) is a minimum. That is,

$$\left(\frac{\partial J}{\partial \alpha}\right)_{\alpha_0} = \left(\frac{\partial J}{\partial \beta}\right)_{\beta_0} = \ldots = 0. \tag{8.78}$$

In general, the accuracy of the calculations can be increased by increasing the number of parameters, eventhough reasonably good results can be obtained with only one parameter if the trial function is chosen judiciously.

We will now apply the method to a few cases by way of illustration:

Linear Harmonic Oscillator

The exact solution of the problem was discussed in Section 4.2A. The Hamiltonian is given (in the co-ordinate representation) by,

$$\hat{H} = -\frac{\hbar^2}{2\mu}\left(\frac{d^2}{dx^2}\right) + \frac{1}{2}Kx^2, \tag{8.79}$$

and the eigenfunctions by Eq. (4.50b).

Let us try two different functions as trial functions. In choosing these we keep in mind that the wavefunction should vanish for $x \to \pm\infty$. Moreover, the probability density should be symmetric about $x = 0$. These conditions are satisfied by

$$\psi_0(x;\alpha) = Ae^{-\frac{1}{2}\alpha x^2}, \tag{8.80a}$$

and

$$\psi_1(x;\beta) = Bxe^{-\frac{1}{2}\beta x^2}. \tag{8.80b}$$

Normalization requires,

$$A^2 = (\alpha/\pi)^{1/2};$$

$$B^2 = 2(\beta^3/\pi)^{1/2}.$$

We have,

$$J_0(\alpha) = \int_{-\infty}^{+\infty} \psi_0 \hat{H} \psi_0 dx = \frac{1}{4\mu\alpha}(\alpha^2\hbar^2 + \mu^2\omega^2),$$

and

$$J_1(\beta) = \int_{-\infty}^{+\infty} \psi_1 \hat{H} \psi_1 dx = \frac{3}{4\mu\beta}(\beta^2\hbar^2 + \mu^2\omega^2),$$

so that,

$$\alpha_0 = (\mu\omega)/\hbar;$$

$$E_0 = J_0(\alpha_0) = \frac{1}{2}\hbar\omega; \tag{8.81a}$$

$$\beta_0 = (\mu\omega)/\hbar;$$

$$E_1 = J_1(\beta_0) = (3/2)\hbar\omega; \tag{8.81b}$$

where,

$$\omega = \sqrt{K/\mu}.$$

We see, from Eq. (4.51), that E_0 and E_1 are, respectively, the energies of the ground state and the first excited state of the oscillator. This result is a consequence of the fact that the trial functions ψ_0 and ψ_1 are exactly of the same form as the actual eigenfunctions of $\hat{H}$ corresponding to the eigenvalues E_0 and E_1. The result emphasizes the fact that a variational calculation does not necessarily lead to the ground state. It is important, therefore, to exercise care in choosing the trial function. For example, if we had made use of the fact that the ground state wavefunction of the oscillator should have no nodes, we would not have chosen ψ_1 (which has a node at $x = 0$) as the trial function in this case. But even with this knowledge, we could choose,

$$\psi_0'(x;\alpha) = C\left(\frac{1}{\alpha^2 + x^2}\right). \tag{8.80c}$$

as the trial function. We would get[15],

$$C^2 = (2\alpha^3/\pi);$$

$$J_0'(\alpha) = \frac{1}{2\mu}(\mu^2\omega^2\alpha^2 + \hbar^2/2\alpha^2),$$

$$\alpha_0^2 = (\sqrt{2}/2)(\hbar/\mu\omega),$$

and

$$E_0' = J_0'(\alpha_0) = \sqrt{2}\left(\frac{1}{2}\hbar\omega\right). \tag{8.81c}$$

Thus, the accuracy is very much reduced when the form of the trial function differs from the actual wave function.

Helium Atom

As another example, let us calculate the ground state energy of the Helium atom. This example will help to compare the variational method with another approximation method—the stationary perturbation theory—discussed in the next section.

The He atom consists of two electrons outside a nucleus consisting of two protons (and one or more neutrons). Hence, the Hamiltonian is given (in the coordinate representation) by,

$$\hat{H} = \hat{H}_1 + \hat{H}_2 + \hat{V}_{12}, \tag{8.82}$$

where,

$$\hat{H}_i = -\frac{\hbar^2}{2\mu}\nabla_i^2 - \frac{Ze^2}{r_i}, \tag{8.82a}$$

and

$$\hat{V}_{12} = \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{e^2}{r_{12}}, \tag{8.82b}$$

$\mathbf{r}_i$ being the position vector of the ith electron (measured from the centre of the nucleus) and Z is the atomic number ($Z = 2$). $\hat{H}_i$ differs from the Hamiltonian of the hydrogen atom [See Eq. (4.97d)] only in the value of the reduced mass μ and

15. Using the following formulae:

$$\int_{-\infty}^{\infty} \frac{dx}{(a^2+x^2)^n} = \frac{(2n-3)!!}{(2n-2)!!}\left(\frac{\pi}{a^{2n-1}}\right), \quad [a > 0;\ n = 2, 3\ldots].$$

$$\int_{-\infty}^{\infty} \frac{x^2}{(\alpha^2+x^2)^2}dx = \frac{2}{\alpha}\int_0^{\pi/2} \sin^2\theta d\theta = \frac{\pi}{2\alpha};$$

$$\int_{-\infty}^{\infty} \frac{x^2}{(\alpha^2+x^2)^4}dx = \frac{2}{\alpha^5}\int_0^{\pi/2} \sin^2\theta\cos^4\theta d\theta = \frac{\pi}{16\alpha^5}.$$

the atomic number Z. Hence, if we neglect the mutual repulsion of the electrons ($\hat{V}_{12}$), then the wavefunctions of the He atom would be merely products of hydrogen-like wavefunctions of the two electrons. In particular, the ground state wavefunction is given by

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \Phi_0(\mathbf{r}_1)\Phi_0(\mathbf{r}_2), \tag{8.83}$$

where, from Eqs. (4.110a), (4.125) and (5.63), we have,

$$\Phi_0(\mathbf{r}) \equiv u_{100}(r, \theta, \phi) = R_{10}(r)Y_{00}(\theta, \phi)$$

$$= (1/\pi a^3)^{1/2} e^{-r/a}, \tag{8.84}$$

where,

$$a = (a_0/Z),$$

and

$$a_0 = (\hbar^2/\mu e^2). \tag{8.85}$$

Thus,

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\pi a^3} e^{-(1/a)(r_1 + r_2)}. \tag{8.83a}$$

ψ given by Eq. (8.83a) can be treated as a trial function for the Hamiltonian $\hat{H}$ given by Eq. (8.82), with Z (or, a) as the variational parameter. The basis for treating Z as the variational parameter is the following: The presence of one electron in the atom partly shields the other electron from the nucleus (by getting in between the electron and the nucleus), thus reducing the effective value of Z.

Since ψ is normalized, we have from Eq. (8.76),

$$E(\psi) \equiv J(Z) = \langle \psi \mid \hat{H} \mid \psi \rangle$$

$$= \langle \Phi_0(\mathbf{r}_1) \mid \hat{H}_1 \mid \Phi_0(\mathbf{r}_1) \rangle + \langle \Phi_0(\mathbf{r}_2) \mid \hat{H}_2 \mid \Phi_0(\mathbf{r}_2) \rangle$$

$$+ \langle \psi \mid \hat{V}_{12} \mid \psi \rangle. \tag{8.86'}$$

Now, from Eq. (4.117) and Problem 4.10, we have, for the case of a hydrogen-like atom,

$$\langle \Phi_0 \mid \hat{T} \mid \Phi_0 \rangle = \frac{Z^2 e^2}{2a_0}, \tag{8.87a}$$

$$\langle \Phi_0 \mid \hat{V} \mid \Phi_0 \rangle = -\frac{Z^2 e^2}{a_0}, \tag{8.87b}$$

where, $\hat{T}$ and $\hat{V}$ are the operators corresponding to the kinetic and the potential energies, respectively. One of the Z in the factor Z^2 in (8.87b) comes from the factor Z in the $\hat{V}$ [see Eq. (8.82a)] while the other Z is due to the Z in the wavefunction (8.84). Now, in the variational method, only the wavefunction is varied, and not the Hamiltonian. Therefore, in (8.82a), we have to use the exact value Z = 2, so that, in place of (8.87b), we have,

$$\langle \Phi_0(\mathbf{r}_i) \mid \hat{V}(r_i) \mid \Phi_0(\mathbf{r}_i) \rangle = -\frac{2Ze^2}{a_0}, \tag{8.87b'}$$

and

$$\langle \Phi_0(\mathbf{r}_i) \mid \hat{H}_i \mid \Phi_0(\mathbf{r}_i) \rangle = (e^2/2a_0)(Z^2 - 4Z), \ (i = 1, 2). \tag{8.88}$$

The last term in (8.86[1]) can be evaluated by writing,

$$\frac{1}{r_{12}} \equiv \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{1}{r_1} \sum_{l=0}^{\infty} (r_2/r_1)^l P_l(\cos\theta), \ r_1 > r_2, \tag{8.89a}$$

$$= \frac{1}{r_2} \sum_{l=0}^{\infty} (r_1/r_2)^l P_l(\cos\theta), \ r_2 > r_1, \tag{8.89b}$$

and using the relationship, (5.152) and Prob. (5.22), where, θ is the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$, we get,

$$\left\langle \psi \left| \frac{e^2}{r_{12}} \right| \psi \right\rangle = e^2 (4\pi)^2 (1/\pi a^3)^2 \int_0^{\infty} \left\{ \int_0^{r_1} (1/r_1) \exp\left[-(2/a)(r_1 + r_2)\right] r_2^2 dr_2 \right.$$

$$\left. + \int_{r_1}^{\infty} (1/r_2) \exp\left[-(2/a)(r_1 + r_2)\right] r_2^2 dr_2 \right\} r_1^2 dr_1$$

$$= \frac{5}{8} \left(\frac{Ze^2}{a_0} \right) \tag{8.90}$$

Substituting from (8.88) and (8.90) in (8.86[1]), we have,

$$J(Z) = (e^2/a_0)[Z^2 - (27/8)Z]. \tag{8.86[2]}$$

Then,

$$Z_0 = \frac{27}{16}, \tag{8.91}$$

and

$$E_0 = J(Z_0) = -2.85(e^2/a_0). \tag{8.92}$$

The result of the first order perturbation theory [Eq. (8.124b)] is obtained by substituting $Z = 2$ in (8.86[2]). Thus,

$$E_0 \text{ (Perturbation)} = -2.75(e^2/a_0) \tag{8.93}$$

The experimental value is, E_0 (experiment) $= -2.904(e^2/a_0)$. We see that, with the same amount of labour, the variational method yields a much better approximation to the ground state energy of the He atom than the perturbation theory.

Excited States

The variational method can be used to get an upper limit to one of the higher energy levels of the system if the trial function is chosen so as to be orthogonal to the wavefunctions of all the lower levels. Suppose $\{\Phi_k\}$ represent the normalized eigenvectors of $\hat{H}$. Then the function,

$$|\psi^n\rangle = |\psi\rangle - \sum_{i=1}^{n-1} |\Phi_i\rangle\langle\Phi_i|\psi\rangle, \tag{8.94}$$

is easily shown to be orthogonal to the $(n-1)$ eigenvectors, Φ_i $(i = 1, 2, \ldots, n-1)$, of $\hat{H}$. If the labelling is done such that $E_{i+1} > E_i$, then according to Eq. (8.75), we have (assuming $|\psi^n\rangle$ to be normalized),

$$E(\psi^n) \equiv \langle\psi^n|\hat{H}|\psi^n\rangle \geq E_n. \tag{8.95}$$

ψ^n is, thus, a trial function for a variational calculation of the nth energy level. An example is provided by the trial function ψ_1 in Eq. (8.80b), which can be verified to be orthogonal to the (ground state) trial function ψ_0 in (8.80a). Therefore, ψ_1 should give an upper limit to the energy of the first excited state of the oscillator. In fact, it leads to the exact energy [Eq. (8.81b)].

Since the error in the determination of the wavefunctions of the lower states would be carried over to that of the higher states, the above method is not practical but for the lowest two or three states. There is another method which is free from this defect: Suppose $\hat{\Omega}$ is an operator (corresponding to an observable) that commutes with $\hat{H}$ and one whose eigenvectors are known. Then a trial function that is constructed entirely from eigenfunctions that belong to a particular eigenvalue of $\hat{\Omega}$, would be orthogonal to all eigenfunctions that belong to the other eigenvalues of $\hat{\Omega}$. A variational calculation with such a trial function would give an upper limit to the lowest energy corresponding to the particular eigenvalue of $\hat{\Omega}$. For example, if $\hat{H}$ has rotational symmetry, $\hat{\Omega}$ could be the angular momentum operator $\hat{\mathbf{L}}$, whose eigenvectors are the spherical harmonics $Y_{lm}(\theta, \phi)$ [See Eq. (5.46)]. Then, a trial function of the type,

$$\psi(\mathbf{r}; \alpha, \beta, \ldots) = f(r; \alpha, \beta, \ldots)Y_{lm}(\theta, \phi), \tag{8.96}$$

will give an upper limit to the lowest energy level with angular momentum l.

Problem 8.5: Use the trial function $\psi_{2p}(\mathbf{r}; \alpha) = A(r/a)\exp[-\alpha(r/a)]Y_{1m}(\theta, \phi)$, to obtain a value for the energy of the $2p$ level in the hydrogen atom. Compare the result with the exact value (4.117).

8.2B Schwinger's Method for Phase Shifts

As an illustration of the application of the variational method to the problem of scattering, let us consider Schwinger's method[16] for the determination of the scattering phase shift.

According to Eq. (7.45[1]), the phase shift δ_l corresponding to the lth partial wave (when the potential is central) is given by

$$\sin\delta_l = -(k/C_l)\int_0^\infty j_l(kr)u_{k,l}(r)U(r)r\,dr, \tag{8.97}$$

16. See, Wu, T.Y. and Ohmura, T., *Quantum Theory of Scattering* (Prentice-Hall, New Jersey 1962), Section D. 2.

where, C_l is given by Eq. (7.34) and $u_{k,l}(r) = rR_{k,l}(r)$, is normalized according to Eq. (7.30) [That is, $\{u_{k,l}(r)/C_l\} \underset{r\to\infty}{\sim} (1/k)\sin(kr - l\pi/2 + \delta_l)$]. The Born approximation replaces $u_{k,l}(r)/C_l$ in Eq. (8.97) by $rj_l(kr)$ [see, Eqs. (7.45[2]) and (7.110a)]. We can obtain a better result by using the variational method. Expression (8.97) is not, however, suited for the application of this method because of the following defects:

(i) It is not stationary since a first order variation in $u_{k,l}(r)$ produces a first order variation in δ_l. This defect can be remedied by making $\sin\delta_l$ quadratic in $u_{k,l}(r)$ just as $E(\psi)$ in (8.69) is quadratic in ψ.

(ii) It depends on the normalization factor C_l which itself depends on δ_l [Eq. (7.34)]. This problem is solved by making Eq. (8.97) homogeneous in $u_{k,l}(r)$.

The objective of making Eq (8.97) both quadratic and homogeneous in $u_{k,l}(r)$ can be achieved with the help of the integral equation for the radial wave function $R_{k,l}(r)$. This integral equation can be obtained from Eq. (7.98) by substituting for $\exp(i\mathbf{k}.\mathbf{r})$, $\psi_k(\mathbf{r})$ and $G_k(\mathbf{r}, \mathbf{r}')$ respectively, from Eqs. (7.24), (7.33a) and (7.100[2]) along with Eq. (E.76). However, the following would be a shorter procedure:

From Eqs. (7.30) and (7.31), we have,

$$R_{k,l}(r) \underset{r\to\infty}{\sim} C_l[\cos\delta_l j_l(kr) - \sin\delta_l n_l(kr)]. \tag{8.98[1]}$$

Substituting for $\sin\delta_l$ here from Eq. (8.97), we get,

$$R_{k,l}(r) \underset{r\to\infty}{\sim} C_l\cos\delta_l j_l(kr) + kn_l(kr)\int_0^\infty j_l(kr')\cdot U(r')R_{k,l}(r')r'^2dr'. \tag{8.98[2]}$$

But, according to Eq. (7.98), we should have (Problem 8.6),

$$R_{k,l}(r) = C_l\cos\delta_l j_l(kr) - \int_0^\infty G_{k,l}(r,r')U(r')R_{k,l}(r')r'^2dr', \tag{8.99a}$$

where, the *spherical Green's function* $G_{k,l}(r,r')$ is defined by

$$G_k(\mathbf{r}, \mathbf{r}') = \sum_{l=0}^{\infty} a_l G_{k,l}(r,r')P_l(\cos\Theta), \tag{8.100}$$

Θ being the angle between $\mathbf{r}$ and $\mathbf{r}'$.

Comparison of Eq. (8.99a) with (8.98[2]) yields,

$$G_{k,l}(r,r') = -kn_l(kr)j_l(kr'),\ r > r'. \tag{8.101a}$$

But, we see from Eq. (7.100[2]) that $G_{k,l}(r,r') = G_{k,l}(r',r)$, so that,

$$G_{k,l}(r,r')_{r<r'} = G_{k,l}(r',r)_{r'>r} = -kn_l(kr')j_l(kr). \tag{8.101b}$$

Thus,

$$R_{k,l}(r) = C_l\cos\delta_l j_l(kr) + kn_l(kr)\int_0^r j_l(kr')U(r')\cdot R_{k,l}(r')r'^2dr'$$

$$+kj_l(kr)\int_r^\infty n_l(kr')U(r')R_{k,l}(r')r'^2dr'; \tag{8.99b}$$

and

$$j_l(kr) = \frac{1}{C_l \cos\delta_l}\left[R_{k,l}(r) + \int_0^\infty G_{k,l}(r,r')U(r')R_{k,l}(r')r'^2 dr'\right]. \qquad (8.102)$$

Eq. (8.99a) or (8.99b) is the integral equation for $R_{k,l}(r)$.

Multiplying both sides of Eq. (8.97) with

$$\int_0^\infty j_l(kr)U(r)R_{k,l}(r)r^2 dr,$$

and then substituting for $j_l(kr)$ on the L.H.S. from (8.102), we get,

$$\frac{\sin\delta_l}{C_l\cos\delta_l}\left[\int_0^\infty R_{k,l}^2(r)U(r)r^2dr + \int_0^\infty U(r)R_{k,l}(r)r^2dr \times \cdot \int_0^\infty G_{k,l}(r,r')U(r')R_{k,l}(r')r'^2dr'\right]$$

$$= -\frac{k}{C_l}\left[\int_0^\infty j_l(kr)U(r)R_{k,l}(r)r^2dr\right]^2,$$

or,

$$k\cot\delta_l = -\frac{\left\{\int_0^\infty R_{k,l}^2(r)U(r)r^2dr + \int_0^\infty U(r)R_{k,l}(r)r^2dr \times \int_0^\infty G_{k,l}(r,r')U(r')R_{k,l}(r')r'^2dr'\right\}}{\left[\int_0^\infty j_l(kr)U(r)R_{k,l}(r)r^2dr\right]^2} \qquad (8.103a)$$

$$-\left[\int_0^\infty u_{k,l}^2(r)U(r)dr - k\int_0^\infty u_{k,l}(r)U(r) \times\left\{n_l(kr)\int_0^r j_l(kr')u_{k,l}(r')U(r')r'dr' + j_l(kr)\int_0^\infty n_l(kr')u_{k,l}(r')U(r')r'dr'\right\}r\,dr\right]$$

$$= \frac{}{\left[\int_0^\infty j_l(kr)u_{k,l}(r)U(r)r\,dr\right]^2} \qquad (8.103b)$$

Eq. (8.103a) is seen to be both quadratic and homogeneous in $R_{k,l}$. It can be shown (Problem 8.7) that $k\cot\delta_l$ is, indeed, stationary with respect to variations of $R_{k,l}$. In fact, Eq. (8.103a) could be written in the form,

$$k\cot\delta_l = \frac{<u_{k,l}\,|\,\hat{A}\,|\,u_{k,l}>}{<u_{k,l}\,|\,\hat{P}\,|\,u_{k,l}>}, \qquad (8.104)$$

where,

$$\hat{A} \equiv A_{k,l}(r,r') = -U(r)\delta(r-r') - U(r)rG_{k,l}(r,r')r'U(r'), \qquad (8.104a)$$

and

$$\hat{P} = P_{k,l}(r,r') = U(r) r j_l(kr) j_l(kr') r' U(r'). \tag{8.104b}$$

Eq. (8.104) is seen to be just an extension of the Ritz variational expression (8.69). On the basis of (8.104), it has been shown[17] that (8.103a) leads to an upper bound for $k \cot \delta_l$ when the potential is attractive ($U(r) \leq 0$) and to a lower bound when the potential is repulsive.

As an application of formula (8.103), let us consider zero-energy scattering by a potential of the form,

$$U(r) = -U_0, \quad r < r_0$$

$$= 0, \quad r > r_0.$$

Only δ_0 would be nonvanishing. We get from Eqs. (8.103b), and (E. 68a, b),

$$k \cot \delta_0 \approx +U_0 \left[\int_0^{r_0} u_0^2(r) dr - \frac{U_0}{k} \int_0^{r_0} u_0(r) \left\{ \cos kr \cdot \int_0^r u_0(r') \sin kr' dr' \right.\right.$$

$$\left.\left. + \sin kr \int_r^{r_0} u_0(r') \cos kr' dr' \right\} dr \right] / \left\{ \frac{-U_0}{k} \int_r^{r_0} u_0(r) \sin kr \, dr \right\}^2$$

$$\approx + \left[\int_0^{r_0} u_0^2(r) dr - U_0 \int_0^{r_0} u_0(r) \left\{ \int_0^r [(r') r' dr' \right.\right.$$

$$\left.\left. + r \int_r^{r_0} u_0(r') dr \right\} dr \right] / \left[U_0 \left\{ \int_0^{r_0} u_0(r) r \, dr \right\}^2 \right].$$

As a trial function, we choose $u_0(r) = r$ (cf. Eq. (7.57a)).

Then, we get,

$$k \cot \delta_0 = \left(\frac{3}{U_0 r_0^3} - \frac{6}{5 r_0} \right). \tag{8.105}$$

The corresponding result in the Born approximation is given by,

$$\sin \delta_0 = -k \int_0^{r_0} j_0^2(kr) (-U_0) r^2 dr$$

$$= k(U_0 r_0^3/3),$$

or

$$k \cot \delta_0 \approx (k / \sin \delta_0) = (3/U_0 r_0^2); \tag{8.106}$$

whereas a more exact result is given by the formula [see Eq. (7.76[2])],

$$k \cot \delta_0 = -(1/a),$$

---

17. Kato, T. *Progress of Theoretical Physics*, **6**, 295 (1951).

where, the scattering length 'a' is given by (Problem 7.2),

$$a = r_0\left[1 - \frac{\tan\sqrt{U_0}r_0}{\sqrt{U_0}r_0}\right].$$

This gives,

$$k\cot\delta_0 = \left[\frac{3}{U_0 r_0^3} - \frac{6}{5r_0} - \cdots\right]. \tag{8.107}$$

Thus, the variational approach in this case proves to be better than the Born approximation. Also, the variational value is on the upper side of the actual value (8.107). It is easily shown, by reversing the sign of U_0 in (8.105) and (8.107), that the variational value would be lower than the actual value for a repulsive potential. However, it is a drawback of the variational method that it does not provide a means of estimating the error when we really do not know the exact value.

Problem 8.6: Obtain Eq. (8.99a) starting from Eq. (7.98).

Problem 8.7: Show that $k\ \cot\delta_l$ given by Eq. (8.103a) is stationary under the variation, $\delta R_{k,l}(r) = \delta R_{k,l}(r)\delta(r - r_1)$.

8.3 STATIONARY PERTURBATION THEORY

This approximation method is useful for finding the changes in the *discrete* energies and the associated wavefunctions of a system resulting from a small disturbance, or *perturbation*, provided the energies and the wavefunctions of the undisturbed system are known. In this method, usually referred to as the *Rayleigh-Schrödinger perturbation theory*, the changes in the energies and the wavefunctions are expressed as an infinite power series in the perturbation parameter (defined below). The approximation, then, consists in neglecting terms in the infinite series after the first few terms. Approximating the series to the first n terms in the series, gives the *nth order approximation.*

Eventhough, we have talked about 'disturbance or perturbation', variation of the Hamiltonian with time is not implied. Either the disturbance was introduced long time ago so that the system has settled down, or the system under consideration differs very little from a system whose energies and wavefunctions are known.

Let $\hat{H}_0$ and $\hat{H}$ represent, respectively, the Hamiltonians of the unperturbed and the perturbed systems. Then,

$$\hat{H} = \hat{H}_0 + \Delta\hat{H}, \tag{8.108}$$

where, $|\Delta\hat{H}|$ is small compared with $|\hat{H}_0|$. We will write,

$$\Delta\hat{H} = \lambda\hat{V}, \tag{8.109}$$

and call λ the *perturbation parameter* (which can vary continuously in the range 0 to 1) and $\hat{V}$ the *perturbing potential*. The solutions of the eigenvalue equation[18],

$$\hat{H}_0\psi_n^{(0)} = E_n^{(0)}\psi_n^{(0)}, \tag{8.110}$$

are known, whereas the solutions of the equation,

$$\hat{H}\psi_n = E_n\psi_n, \tag{8.111}$$

are to be found.

Let

$$E_n = E_n^{(0)} + \Delta E_n, \tag{8.112a}$$

and

$$\psi_n = \psi_n^{(0)} + \Delta\psi_n. \tag{8.112b}$$

$\Delta\psi_n$ and ΔE_n are small since $|\Delta\hat{H}|$ is small. Substituting in (8.111) from (8.108) and (8.112a, b), we get

$$\hat{H}_0\psi_n^{(0)} + (\Delta\hat{H})\psi_n + \hat{H}_0(\Delta\psi_n) = E_n^{(0)}\psi_n^{(0)} + E_n^{(0)}(\Delta\psi_n) + (\Delta E_n)\psi_n. \tag{8.113}$$

In view of Eq. (8.110), the terms $\hat{H}_0\psi_n^{(0)}$ and $E_n^{(0)}\psi_n^{(0)}$ in (8.113) cancel each other. Then, taking the scalar product of the latter equation with $\psi_n^{(0)}$, we get,

$$\langle\psi_n^{(0)} | \Delta\hat{H} | \psi_n\rangle + \langle\psi_n^{(0)} | \hat{H}_0 | \Delta\psi_n\rangle$$

$$= E_n^{(0)}\langle\psi_n^{(0)} | \Delta\psi_n\rangle + \Delta E_n\langle\psi_n^{(0)} | \psi_n\rangle.$$

But,

$$\langle\psi_n^{(0)} | \hat{H}_0 | \Delta\psi_n\rangle = \langle\hat{H}_0\psi_n^{(0)} | \Delta\psi_n\rangle = E_n^{(0)}\langle\psi_n^{(0)} | \Delta\psi_n\rangle$$

Therefore,

$$\Delta E_n = \frac{\langle\psi_n^{(0)} | \Delta\hat{H} | \psi_n\rangle}{\langle\psi_n^{(0)} | \psi_n\rangle} = \lambda\frac{\langle\psi_n^{(0)} | \hat{V} | \psi_n\rangle}{\langle\psi_n^{(0)} | \psi_n\rangle}. \tag{8.114}$$

The method consists in writing,

$$\Delta\psi_n = \lambda\psi_n^{(1)} + \lambda^2\psi_n^{(2)} + \ldots = \sum_{s=1}^{\infty}\lambda^s\psi_n^{(s)}. \tag{8.115a}$$

Also, we assume that

$$\langle\psi_n^{(0)} | \Delta\psi_n\rangle = 0, \tag{8.116a}$$

so that,

$$\langle\psi_n^{(0)} | \psi_n\rangle = \langle\psi_n^{(0)} | \psi_n^{(0)}\rangle = 1. \tag{8.117}$$

The assumption (8.116a) merely represents a particular choice for the normalization of ψ_n. For, ψ_n can be expanded in terms of the complete, orthonormal set $\{\psi_n^{(0)}\} \equiv \{\mathbf{u}_n\}$, of eigenvectors of $\hat{H}_0$:

---

18. We assume that the spectrum of $\hat{H}_0$ is *discrete*.

$$\psi_n = \sum_k c_k \mathbf{u}_k = \mathbf{u}_n + \sum_{k \neq n} c_k \mathbf{u}_k + (c_n - 1)\mathbf{u}_n.$$

Thus,

$$\Delta\psi_n = (c_n - 1)\mathbf{u}_n + \sum_{k \neq n} c_k \mathbf{u}_k,$$

and

$$\langle \psi_n^{(0)} \mid \Delta\psi_n \rangle \equiv \langle \mathbf{u}_n \mid \Delta\psi_n \rangle = (c_n - 1).$$

Eq. (8.116a) requires that $c_n = 1$, so that,

$$\langle \psi_n \mid \psi_n \rangle = \sum_k |c_k|^2 = 1 + \sum_{k \neq n} |c_k|^2 > 1.$$

From Eqs. (8.115a) and (8.116a), we have,

$$\langle \psi_n^{(0)} \mid \psi_n^{(s)} \rangle = 0, \quad s \geq 1. \tag{8.116b}$$

Substituting from Eqs. (8.115a) and (8.117) in (8.114), we get

$$\Delta E_n = \sum_{s=1}^{\infty} \lambda^s E_n^{(s)}, \tag{8.115b}$$

where,

$$E_n^{(s)} = \langle \psi_n^{(0)} \mid \hat{V} \mid \psi_n^{(s-1)} \rangle, \; s \geq 1. \tag{8.118}$$

Thus, a knowledge of the wave function to a particular order enables us to calculate the energy to the next higher order. And the wavefunction to a particular order is determined from Eq. (8.111) which, according to Eqs. (8.108), (8.109), (8.112a, b) and (8.115a, b) becomes,

$$(\hat{H}_0 + \lambda\hat{V})(\psi_n^{(0)} + \lambda\psi_n^{(1)} + \ldots)$$

$$= (E_n^{(0)} + \lambda E_n^{(1)} + \ldots) \cdot (\psi_n^{(0)} + \lambda\psi_n^{(1)} + \ldots). \tag{8.119}$$

Since the series (8.115a, b) are assumed to be continuous, analytic functions of λ for $0 \leq \lambda \leq 1$, coefficients of like powers of λ on either side of Eq. (8.119) should be equal. Hence we have,

$$\lambda^0 : (\hat{H}_0 - E_n^{(0)})\psi_n^{(0)} = 0, \tag{8.120^0}$$

$$\lambda^1 : (\hat{H}_0 - E_n^{(0)})\psi_n^{(1)} = (E_n^{(1)} - \hat{V})\psi_n^{(0)}, \tag{8.120^1}$$

$$\lambda^2 : (\hat{H}_0 - E_n^{(0)})\psi_n^{(2)} = E_n^{(2)}\psi_n^{(0)} + (E_n^{(1)} - \hat{V})\psi_n^{(1)}, \tag{8.120^2}$$

.....

$$\lambda^S : (\hat{H}_0 - E_n^{(0)})\psi_n^{(s)} = E_n^{(s)}\psi_n^{(0)} + E_n^{(s-1)}\psi_n^{(1)} + \ldots + E_n^{(2)}\psi_n^{(s-2)} + (E_n^{(1)} - \hat{V})\psi_n^{(s-1)}. \tag{8.120^s}$$

At this stage, we have to make a distinction between degenerate and non-degenerate case

8.3A. Nondegenerate Case

Taking the scalar product of (8.120[s]) with $\psi_k^{(0)} \neq \psi_n^{(0)}$, we get,

$$(E_k^{(0)} - E_n^{(0)}) \langle \psi_k^{(0)} \mid \psi_n^{(s)} \rangle = \langle \psi_k^{(0)} \mid \{(E_n^{(1)} - \hat{V}) \mid \psi_n^{(s-1)}\rangle + E_n^{(2)} \mid \psi_n^{(s-2)}\rangle$$

$$+ \ldots + E_n^{(s-1)} \mid \psi_n^{(1)}\rangle\}, \quad (8.121a)$$

or,

$$\sum_{k \neq n} \mid \psi_k^{(0)}\rangle \langle \psi_k^{(0)} \mid \psi_n^{(s)}\rangle = \sum_{k \neq n} \frac{\mid \psi_k^{(0)} > < \psi_k^{(0)} \mid}{E_k^{(0)} - E_n^{(0)}}$$

$$\cdot \{(E_n^{(1)} - \hat{V}) \mid \psi_n^{(s-1)}\rangle + E_n^{(2)} \mid \psi_n^{(s-2)}\rangle + \ldots + E_n^{(s-1)} \mid \psi_n^{(1)}\rangle\} \quad (8.121b)$$

But $\sum_k \mid \psi_k^{(0)}\rangle \langle \psi_k^{(0)} \mid \equiv \sum_k \mid \mathbf{u}_k\rangle \langle \mathbf{u}_k \mid = \hat{1}$,

so that,

$$\sum_{k \neq n} \mid \psi_k^{(0)}\rangle \langle \psi_k^{(0)} \mid = \hat{1} - \mid \psi_n^{(0)}\rangle \langle \psi_n^{(0)} \mid . \quad (8.122)$$

Substituting from (8.122) and (8.116b) on the L.H.S. of (8.121b), we have,

$$\mid \psi_n^{(s)}\rangle = \sum_{k \neq n} \frac{\mid \psi_k^{(0)} > < \psi_k^{(0)} \mid}{E_k^{(0)} - E_n^{(0)}} \{(E_n^{(1)} - \hat{V}) \mid \psi_n^{(s-1)}\rangle + \ldots$$

$$+ E_n^{(s-1)} \mid \psi_n^{(1)}\rangle\}. \quad (8.123)$$

This is a recurrence relation for $\psi_n^{(s)}$. Together with Eq. (8.118), it enables us to determine successively $\psi_n^{(1)}, \psi_n^{(2)}, \ldots$, starting from $\psi_n^{(0)}$. Thus, putting $s = 1$ in (8.123), we get (since $\langle \psi_k^{(0)} \mid \psi_n^{(0)}\rangle = 0$),

$$\psi_n^{(1)} = \sum_{k \neq n} \frac{\mid \psi_k^{(0)} > < \psi_k^{(0)} \mid \hat{V} \mid \psi_n^{(0)} >}{E_n^{(0)} - E_k^{(0)}}$$

$$= \sum_{k \neq n} \left\{ \frac{V_{kn}}{E_n^{(0)} - E_k^{(0)}} \right\} \psi_k^{(0)}, \quad (8.124a)$$

where, V_{kn} is the knth matrix element of V in the representation spanned by the basis vectors $\{\mathbf{u}_k\}$. Similarly, from Eq. (8.118), we have

$$E_n^{(1)} = < \psi_n^{(0)} \mid \hat{V} \mid \psi_n^{(0)} > = V_{nn}. \quad (8.124b)$$

The wavefunctions and the energies up to the *first order in the perturbation*, are, then, given by

$$\psi_n^I = \psi_n^{(0)} + \lambda \psi_n^{(1)} = \psi_n^{(0)} + \lambda \sum_{k \neq n} \frac{V_{kn}}{(E_n^{(0)} - E_k^{(0)})} \psi_k^{(0)}, \quad (8.125a)$$

and

$$E_n^I = E_n^{(0)} + \lambda V_{nn}. \quad (8.125b)$$

Similarly, putting $s = 2$ in Eqs. (8.123) and (8.118) and substituting for $\psi_n^{(1)}$ and $E_r^{(1)}$ from (8.124a, b), we get,

$$\psi_n^{(2)} = \sum_{k \neq n} \frac{|\psi_k^{(0)}><\psi_k^{(0)}|(E_n^{(1)} - \hat{V})|\psi_n^{(1)}>}{E_k^{(0)} - E_n^{(0)}}$$

$$= \sum_{k \neq n} \left\{ \sum_{m \neq n} \frac{V_{km} V_{mn}}{(E_n^{(0)} - E_m^{(0)})(E_n^{(0)} - E_k^{(0)})} - \frac{V_{kn} V_{nn}}{(E_n^{(0)} - E_k^{(0)})^2} \right\} \psi_k^{(0)}, \qquad (8.126a)$$

$$E_n^{(2)} = <\psi_n^{(0)}|\hat{V}|\psi_n^{(1)}> = \sum_{k \neq n} \frac{|V_{kn}|^2}{E_n^{(0)} - E_k^{(0)}}, \qquad (8.126b)$$

Eqs. (8.126a, b) represent the *second-order contribution* (when multiplied by λ^2) to the wavefunction and the energy.[19]

ILLUSTRATIVE EXAMPLES

Anharmonic Oscillator

Consider a linear harmonic oscillator subjected to a small force represented by the potential $V(x) = Cx^4$. Then [(see section 4.2A)]

$$\hat{H}_0 = (\hat{p}^2/2\mu) + (1/2)K\hat{x}^2$$

$$\Delta\hat{H} = \hat{V}(x) = C\hat{x}^4. \text{ (That is, } \lambda = 1)$$

$\psi_n^{(0)} \equiv u_n(x)$ and $E_n^{(0)}$ are given, respectively by Eqs. (4.50b) and (4.51). From Eq. (8.124b), we have,

$$E_n^{(1)} = C <u_n|\hat{x}^4|u_n>$$

$$= C\sum_m <u_n|\hat{x}^2|u_m><u_m|\hat{x}^2|u_n>.$$

From Eqs. (4.47a), (4.50a, b) and (E.12b), we get,

$$\langle u_m|\hat{x}^2|u_n\rangle = \frac{1}{\alpha^2}\int_{-\infty}^{+\infty} \phi_m^*(\xi)\xi^2\phi_n(\xi)d\xi$$

$$= \frac{1}{2\alpha^2}[\sqrt{n(n-1)}\,\delta_{m,n-2} + (2n+1)\delta_{m,n}$$

$$+ \sqrt{(n+1)(n+2)}\,\delta_{m,n+2}], \qquad (8.127)$$

with $\alpha^2 = (K/\hbar\omega)$: $\omega = \sqrt{K/\mu}$. Hence,

$$E_n^{(1)} = C\sum_m |\langle u_m|\hat{x}^2|u_n\rangle|^2 = \frac{3C}{3K^2}\left\{\left(n + \frac{1}{2}\right)^2 + \frac{1}{4}\right\}\hbar^2\omega^2;$$

---

19. Customarily, λ is set equal to 1 in the final expressions for ψ_n and E_n.

and, up to the first order in the perturbation, the energy of the oscillator is given by

$$E_n^I = \left(n+\frac{1}{2}\right)\hbar\omega + \frac{3C}{2K^2}\left\{\left(n+\frac{1}{2}\right)^2+\frac{1}{4}\right\}(\hbar\omega)^2,\ n=0,1,2,\ldots+\infty. \tag{8.128a}$$

The wavefunction to the same order is given by

$$\psi_n^I = \mathbf{u}_n + C\sum_{k\neq n}\left\{\frac{\langle u_k \mid \hat{x}^4 \mid u_n\rangle}{E_n^{(0)}-E_k^{(0)}}\right\}\mathbf{u}_k$$

$$= \mathbf{u}_n + C\left\{\frac{1}{4\hbar\omega}\langle u_{n-4}\mid \hat{x}^4\mid u_n\rangle \mathbf{u}_{n-4} + \frac{1}{2\hbar\omega}\langle u_{n-2}\mid \hat{x}^4\mid u_n\rangle \mathbf{u}_{n-2}\right.$$

$$\left. -\frac{1}{2\hbar\omega}\langle u_{n+2}\mid \hat{x}^4\mid u_n\rangle \mathbf{u}_{n+2} - \frac{1}{4\hbar\omega}\langle u_{n+4}\mid \hat{x}^4\mid u_n\rangle \mathbf{u}_{n+4}\right\}$$

$$= \mathbf{u}_n(x) + \frac{C\hbar\omega}{K^2}\left\{\frac{1}{2^4}\sqrt{\frac{n!}{(n-4)!}}\mathbf{u}_{n-4} + \frac{2n-1}{2^2}\sqrt{\frac{n!}{(n-2)!}}\mathbf{u}_{n-2}\right.$$

$$\left. -\frac{2n+3}{2^2}\sqrt{\frac{(n+2)!}{n!}}\mathbf{u}_{n+2} - \frac{1}{2^4}\sqrt{\frac{(n+4)!}{n!}}\mathbf{u}_{n+4}\right\}, \tag{8.128b}$$

where, $\mathbf{u}_n(x)$ is given by Eq. (4.50b).

Linear Harmonic Oscillator

As another example, let us consider a linear harmonic oscillator under the influence of a perturbing field represented by $V(x)=\frac{1}{2}bx^2$. The harmonic nature of the oscillator would, obviously, be preserved by such a field. From formulae (8.124 a, b) and (8.126a, b) and using Eq. (8.127), we get,

$$E_n^{(1)} = (b/2K)\left(n+\frac{1}{2}\right)\hbar\omega,$$

$$E_n^{(2)} = -(b^2/8K^2)\left(n+\frac{1}{2}\right)\hbar\omega,$$

$$\psi_n^{(1)} = (b/8K)\{\sqrt{n(n-1)}\,\mathbf{u}_{n-2} - \sqrt{(n+1)(n+2)}\,\mathbf{u}_{n+2}\},$$

$$\psi_n^{(2)} = (b^2/16K^2)\{\sqrt{(n+1)(n+2)}\,\mathbf{u}_{n+2} - \sqrt{n(n-1)}\,\mathbf{u}_{n-2}\},$$

so that, up to the second order in perturbation, the energy and the wavefunction of the nth oscillator level, are given by

$$E_n^{II} = \left(n+\frac{1}{2}\right)\hbar\omega\left[1+\frac{b}{2K}-\frac{b^2}{8K^2}\right], \tag{8.129a}$$

$$\psi_n^{II} = \mathbf{u}_n(x) + \frac{b}{8K}\left(1-\frac{b}{2K}\right)\{\sqrt{n(n-1)}\,\mathbf{u}_{n-2} - \sqrt{(n+1)(n+2)}\,\mathbf{u}_{n+2}\}. \tag{8.129b}$$

In this case, however, the problem is capable of an exact solution as the Hamiltonian is given by

$$\hat{H} = \frac{\hat{p}^2}{2\mu} + \frac{1}{2}K'\hat{x}^2,$$

with $K' = K + b$. Then, from Eqs. (4.51), and (4.47b), we have,

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega' = \left(n + \frac{1}{2}\right)\hbar\sqrt{\frac{K+b}{\mu}}$$

$$= \left(n + \frac{1}{2}\right)\hbar\omega\left\{1 + \frac{b}{2K} - \frac{b^2}{8K^2} + \frac{1}{16}\frac{b^3}{K^3} - \ldots\right\}, \tag{8.130a}$$

and from Eq. (4.50a, b), we get,

$$\psi_n(x) = \sqrt{\alpha'}\,\phi_n(\xi')$$

$$\approx \left(1 + \frac{1}{8}\frac{b}{K} - \frac{7}{128}\frac{b^2}{K^2}\right)\sqrt{\alpha}\,\phi_n\left(\xi + \frac{b}{4K}\xi - \frac{3}{32}\frac{b^2}{K^2}\xi^2\right), \tag{8.131}$$

where

$$\alpha' = \left(\frac{\mu(K+b)}{\hbar^2}\right)^{1/4};$$

$$\alpha = \left(\frac{\mu K}{\hbar^2}\right)^{1/4};$$

$$\xi' = \alpha' x;$$

$$\xi = \alpha x.$$

Expanding $\phi_n(\,)$ in Taylor series around ξ, we have,

$$\phi_n\left(\xi + \frac{b}{4K}\xi - \frac{3}{32}\frac{b^2}{K^2}\xi\right)$$

$$\approx \phi_n(\xi) + \frac{b}{4K}\xi\phi_n'(\xi) + \frac{b^2}{32K^2}\{\xi^2\phi_n''(\xi) - 3\xi\phi_n'(\xi)\}, \tag{8.132}$$

where, the prime on ϕ_n denotes differentiation with respect to ξ. Substituting (8.132) in (8.131) and using Eqs. (E. 11a-d) we get,

$$\psi_n(x) \approx \mathbf{u}_n(x) + \frac{b}{8K}\left(1 - \frac{b}{2K}\right)\{\sqrt{n(n-1)}\,\mathbf{u}_{n-2}(x) - \sqrt{(n+1)(n+2)}\,\mathbf{u}_{n+2}(x)\} + \ldots,$$

(8.130b)

where,

$$\mathbf{u}_n(x) = \sqrt{\alpha}\,\phi_n(\alpha x) \equiv \psi_n^{(0)}.$$

Comparison of (8.129a, b) with (8.130a, b) verifies that the perturbation series gives correctly the series expansion of the energy and the wavefunction in powers of the ratio (b/K).

Problem 8.8: A one-dimensional harmonic oscillator is subjected to a constant force F. Calculate the shift in the energy levels and the wave functions up to the second order in F.

Problem 8.9: The Hamiltonian of a hydrogen-like atom in an electric field $\mathbf{E}$ can be written as

$$\hat{H} = \hat{H}_0 - \mathbf{E} \cdot \hat{\mathbf{d}},$$

where

$$\hat{\mathbf{d}} = -e\hat{\mathbf{r}},$$

is the electric dipole moment operator for the atom. Show that, up to second order in the perturbation, the energy of the atom is given by the expression,

$$E_n'' = E_n^{(0)} - (\mathbf{E} \cdot \mathbf{d}_0) - \frac{1}{2}(\mathbf{E} \cdot \mathbf{d}_1),$$

where $\mathbf{d}_0$ and $\mathbf{d}_1$ are, respectively, the *permanent* (independent of $\mathbf{E}$) and the *induced* (proportional to $\mathbf{E}$) dipole moments.

8.3B. Degenerate Case

When $\hat{H}_0$ has degenerate eigenvalues, the above method requires modification as Eq. (8.124a) and hence also the equations of higher order, breaks down when $E_n^{(0)} = E_k^{(0)}$, unless $\langle \mathbf{u}_k \mid \hat{V} \mid \mathbf{u}_n \rangle = 0$. This is because the factor $\langle \mathbf{u}_k \mid \hat{V} \mid \mathbf{u}_n \rangle / (E_n^{(0)} - E_k^{(0)})$ becomes too large for the validity of the perturbation approximation. We can circumvent this difficulty by replacing the zero-order eigenfunctions, $\mathbf{u}_k$, belonging to the degenerate eigenvalues by linear combinations, Φ_k, of these functions such that $\langle \Phi_k \mid \hat{V} \mid \Phi_n \rangle = 0$ for $E_k^{(0)} = E_n^{(0)}$. This procedure is equivalent to diagonalizing $\hat{V}$ in the subspace spanned by the degenerate eigenfunctions of $\hat{H}_0$. The eigenvalues of $\hat{V}$ thus obtained would be the first order corrections $E_n^{(1)}$ to the energy $E_n^{(0)}$, and the eigenfunctions of $\hat{V}$ would be the Φ_n's. We will illustrate the procedure by assuming that $\hat{H}_0$ has an eigenvalue which is g-fold degenerate.

Let

$$E_1^{(0)} = E_2^{(0)} = \ldots = E_g^{(0)} = E^{(0)}, \tag{8.133}$$

represent the degenerate eigenvalues and let the g linearly independent eigenvectors belonging to these eigenvalues be $\mid \mathbf{u}_k \rangle$, $(k = 1, \ldots g)$. Then,

$$\hat{H}_0 \mid \mathbf{u}_k \rangle = E^{(0)} \mid \mathbf{u}_k >, \qquad k \leq g,$$

but,

$$\langle \mathbf{u}_k \mid \hat{V} \mid \mathbf{u}_j \rangle \neq 0, \text{ for } k, j \leq g.$$

Define,

$$\left.\begin{aligned} |\Phi_k\rangle &= \sum_{i=1}^{g} S_{ik}\,|u_i\rangle,\ k = 1,2,\ldots g \\ &= |\mathbf{u}_k\rangle,\ k > g. \end{aligned}\right\} \tag{8.134}$$

Obviously,

$$\hat{H}_0\,|\Phi_k\rangle = E_k^{(0)}\,|\Phi_k\rangle. \tag{8.135}$$

We have to determine the coefficients S_{ik} in (8.134) such that

$$\langle\Phi_k\,|\,\hat{V}\,|\,\Phi_n\rangle = \delta_{kn}E_n^{(1)},\ \text{for } n, k \le g, \tag{8.136}$$

where, we have assumed that the Φ_k's are normalized just as the $\mathbf{u}_k$'s are. The following procedure would accomplish the objective:

Replace $\psi_n^{(0)}$ in (8.120[1]) by Φ_n: We get, using (8.134),

$$(\hat{H}_0 - E_n^{(0)})\psi_n^{(1)} = (E_n^{(1)} - \hat{V})\Phi_n$$

$$= (E_n^{(1)} - \hat{V})\sum_{j=1}^{g} S_{jn}\,|u_j\rangle,\ n \le g. \tag{8.137a}$$

$$= (E_n^{(1)} - \hat{V})\,|u_n\rangle,\ n > g. \tag{8.137b}$$

Taking the scalar product of Eq. (8.137a) with $\mathbf{u}_k (k \le g)$, we get, since $E_k^{(0)} = E_n^{(0)}$,

$$\sum_{j=1}^{g}\{V_{kj} - \delta_{kj}\cdot E_n^{(1)}\}S_{jn} = 0,\ (k = 1,2,\ldots g;\ n \le g),$$

or,

$$(V^g - E^{(1)}I)S_n = 0,\ (n = 1,2,\ldots,g), \tag{8.138}$$

where V^g and S_n are the matrices,

$$V^g = \begin{pmatrix} V_{11} & V_{12} & .. & V_{1g} \\ V_{21} & V_{22} & .. & V_{2g} \\ . & & & \\ V_{g1} & V_{g2} & .. & V_{gg} \end{pmatrix};\ S_n = \begin{pmatrix} S_{1n} \\ S_{2n} \\ . \\ . \\ S_{gn} \end{pmatrix}, \tag{8.139}$$

with

$$V_{kj} \equiv \langle u_k\,|\,\hat{V}\,|\,u_j\rangle. \tag{8.140}$$

Also, I is a $g \times g$ unit matrix. Eq. (8.138) is the eigenvalue equation [see Eq. (A.63)] of the matrix V^g. The eigenvalues are the g roots of the secular equation (cf. Eq. (A. 58)) $|\,V^g - E^{(1)}I\,| = 0$. These eigenvalues, $E_n^{(1)}$ (for $n = 1$ to g), give the first order corrections to the energy $E_n^{(0)}$, while the eigenvectors of V^g (that is, the Φ_n's given by Eq. (8.134) for $n \le g$) are the zero-order wavefunctions for $n \le g$. The problem is, thus, reduced to that of diagonalizing the matrix V^g.

The matrix V_Φ representing $\hat{V}$ in the *representation* $\{\Phi_k\}_N$ is shown below against the matrix V_u representing $\hat{V}$ in the representation $\{\mathbf{u}_k\}_N$, where N is the number of distinct (linearly independent) eigenvectors of $\hat{H}_0$:

·	u_1	u_2	...	u_g	u_{g+1}	...	u_N
u_1	V_{11}	V_{12}	...	V_{1g}	V_{1g+1}	...	V_{1N}
u_2	V_{21}	V_{22}	...	V_{2g}	V_{2g+1}	...	V_{2N}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
u_g	V_{g1}	V_{g2}	...	V_{gg}	V_{gg+1}	...	V_{gN}
u_{g+1}	V_{g+11}	V_{g+12}	...	V_{g+1g}	$V_{g+1\,g+1}$	...	V_{g+1N}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
u_N	V_{N1}	V_{N2}	...	V_{Ng}	$V_{N\,g+1}$	...	V_{NN}

$$= V_u = \begin{pmatrix} V^g & V^{g,(N-g)} \\ V^{(N-g),g} & V^{(N-g)} \end{pmatrix} \tag{8.141a}$$

	Φ_1	Φ_2	...	Φ_g	Φ_{g+1}	...	Φ_N
Φ_1	$E_1^{(1)}$	0	...	0	V'_{1g+1}	...	V'_{1N}
Φ_2	0	$E_2^{(1)}$	...	0	V'_{2g+1}	...	V'_{2N}
.	.	.	...	.	.	...	
.	.	.		.	.		
Φ_g	0	0	...	$E_g^{(1)}$	$V'_{g\,g+1}$	...	V'_{gN}
Φ_{g+1}	V'_{g+11}	V'_{g+12}	...	V'_{g+1g}	$V_{g+1\,g+1}$	...	V_{g+1N}
.	.						
.	.						
Φ_N	V'_{N1}	V'_{N2}	...	V'_{Ng}	$V_{N\,g+1}$	...	V_{NN}

$$= V_\Phi = \begin{pmatrix} E_g & V'^{g,(N-g)} \\ V'^{(N-g),g} & V^{(N-g)} \end{pmatrix} \tag{8.141b}$$

In (8.141b),

$$V'_{ik} = \langle \Phi_i \mid \hat{V} \mid \Phi_k \rangle = \sum_{j=1}^{g} S_{ji}^* V_{jk}, \quad \text{for } i < k,$$

$$= \sum_{j=1}^{g} S_{jk} V_{ij}, \quad \text{for } i > k. \tag{8.142}$$

It follows that

$$\left.\begin{aligned} E_n^{(1)} &= \langle \Phi_n \mid \hat{V} \mid \Phi_n \rangle, \\ \psi_n^{(1)} &= \sum_{k \neq n} \left\{ P \frac{\langle \Phi_k \mid \hat{V} \mid \Phi_n \rangle}{E_n^{(0)} - E_k^{(0)}} \right\} \Phi_k \end{aligned}\right\} \quad n = 1, 2, \ldots N. \tag{8.143}$$

Thus, up to first order in the perturbation, we have,

$$\left.\begin{aligned} E_n' &= E^{(0)} + E_n^{(1)}, \; n \leq g, \\ &= E_n^{(0)} + V_{nn}, \; n > g, \end{aligned}\right\} \tag{8.144a}$$

$$\left.\begin{aligned}\psi_n' &= \Phi_n + \sum_{k>g}\sum_{j=1}^{g}\left\{\frac{V_{kj}S_{jn}}{E_n^{(0)}-E_k^{(0)}}\right\}\mathbf{u}_k,\ n \le g,\\ &= \mathbf{u}_n + \sum_{k\neq n}\left\{\frac{V_{kn}}{E_n^{(0)}-E_k^{(0)}}\right\}\mathbf{u}_k,\ n > g,\end{aligned}\right\} \tag{8.144b}$$

where, we have substituted for Φ_k from Eq. (8.134) in the expression (8.143) for $\psi_n^{(1)}$.

If all the eigenvalues $E_n^{(1)}$ $(n = 1, 2, ..., g)$ of V^g are different, then the degeneracy is completely removed by the perturbation in the first order, and the single level $E^{(0)}$ is split up into g different levels (Fig. 8.6). If some of the eigenvalues are equal, then the degeneracy is removed only partly in the first order. The degeneracy may get removed completely in a higher order; but, sometimes, the degeneracy might be there in all orders of the perturbation.

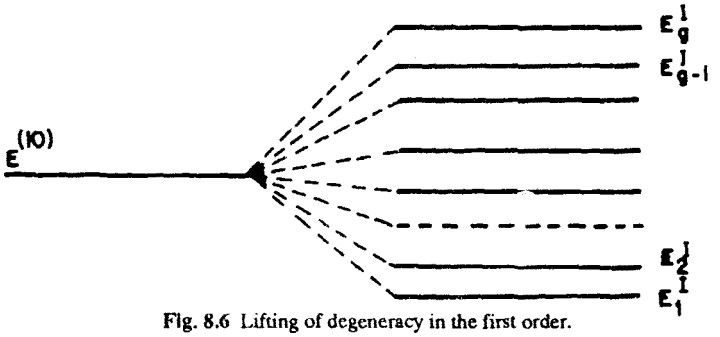


Fig. 8.6 Lifting of degeneracy in the first order.

A sufficient condition for the persistence of the degeneracy in the first order is that two or more of the zero-order degenerate wavefunctions satisfy the relations,

$$< u_i \mid \hat{V} \mid u_i > = < u_j \mid \hat{V} \mid u_j >, \tag{8.145a}$$

$$< u_i \mid \hat{V} \mid u_k > = < u_j \mid \hat{V} \mid u_k > = 0,\ k \le g; \tag{8.145b}$$

For, (8.145b) implies, $u_i \equiv \Phi_i$ and $u_j \equiv \Phi_j$ so that, from Eqs. (8.143, 144 or, 145a), we have, $E_i' = E_j'$. The conditions (8.145a, b) are, however, not necessary except for $g = 2$. Thus, a sufficient condition for the complete removal of the degeneracy in the first order is that the perturbation connect *all* the zero-order wave functions in the first order. In general, degeneracy is removed in the nth order if the perturbation connects all the zero-order wavefunctions in the nth order. That is, if

$$V_{ij}^{(n)} \equiv \sum_{k,l,m} < i \mid \hat{V} \mid k >< k \mid \hat{V} \mid l \times\!< l \mid \hat{V} \mid m > \ldots . \hat{V} \mid j > \neq 0,$$

(n factors)

where, $\quad \mid i > \equiv \mid u_i >.$

APPLICATIONS

Linear Stark Effect in the Hydrogen Atom

Stark Effect is the splitting of atomic levels due to an applied electric field. In the case of the hydrogen atom, the Hamiltonian would be given by

$$\hat{H} = \hat{H}_0 + e \mid \mathbf{E} \mid \hat{z}.$$

where $\mathbf{E}$ is the electric field, assumed to be in the z-direction, and $\hat{H}_0$ is given by Eqs. (4.97d) and (4.98a). Thus, the perturbation,

$$\hat{V} = e \mid \mathbf{E} \mid \hat{z}.$$

The zero-order wavefunctions are given by Eq. (4.110a):

$$\psi_k^{(0)} \equiv \mathbf{u}_k = u_{nlm}(r, \theta, \phi).$$

The ground state [See Eqs. (4.117) and (4.123)] is, u_{100} which is non-degenerate. The first excited state corresponds to $n = 2(l = 0, 1)$ and is 4-fold degenerate, the degenerate eigenvectors being $u_{200}, u_{210}, u_{21-1}$ and u_{211}. Since $\hat{V}$ has odd parity, its matrix elements between states of the same parity (that is, same l-value) vanish.

Thus,

$$< u_{2lm} \mid \hat{V} \mid u_{2lm} > = 0, \tag{8.146a}$$

so that

$$< u_{21-1} \mid \hat{V} \mid u_{21m} > = < u_{211} \mid \hat{V} \mid u_{21m} > = 0. \tag{8.146b}$$

Also, from $z = r\cos\theta = \sqrt{\frac{4\pi}{3}} r Y_{10}(\theta, \phi)$, and the relation (5.60), we have,

$$< u_{21-1} \mid \hat{V} \mid u_{200} > = < u_{211} \mid \hat{V} \mid u_{200} > = 0. \tag{8.146c}$$

We see that u_{21-1} and u_{211}, thus, satisfy conditions (8.145a, b) so that they will continue to be degenerate eigenvectors in the first order also. Hence, we have, according to (8.134),

$$\Phi_1 = S_{11} u_{200} + S_{21} u_{210},$$

$$\Phi_2 = S_{12} u_{200} + S_{22} u_{210}$$

$$\Phi_3 = u_{21-1},$$

$$\Phi_4 = u_{211}.$$

Only the levels u_{200} and u_{210} split, and the splitting is obtained by diagonalizing the matrix,

$$V^{s=2} = \begin{pmatrix} 0 & V_{12} \\ V_{21} & 0 \end{pmatrix},$$

where,

$$V_{12} = < u_{200} \mid \hat{V} \mid u_{210} >$$

$$= e \mid \mathbf{E} \mid < u_{200} \mid \hat{z} \mid u_{210} >$$

$$= e \mid \mathbf{E} \mid \int\int\int u^{*}_{200}\, r \cos\theta\, u_{210} r^2 dr\, d(\cos\theta) d\phi$$

$$= -3e \mid \mathbf{E} \mid a_0,$$

since

$$u_{200} = \frac{1}{\sqrt{4\pi}} \left(\frac{1}{2a_0}\right)^{3/2} \left(2 - \frac{1}{a_0}\right) e^{r/2a_0},$$

$$u_{210} = \frac{1}{\sqrt{4\pi}} \left(\frac{1}{2a_0}\right) (r/a_0) e^{-r/2a_0} \cos\theta,$$

where,

$$a_0 = (\hbar^2/\mu e^2),$$

is the Bohr radius.

$$V_{21} = V^{*}_{12} = V_{12}.$$

Hence, the secular equation is

$$\begin{vmatrix} -E^{(1)} & -3e \mid \mathbf{E} \mid a_0 \\ -3e \mid \mathbf{E} \mid a_0 & -E^{(1)} \end{vmatrix} = 0,$$

yielding the roots,

$$E^{(1)}_1 = -3e \mid \mathbf{E} \mid a_0;$$

$$E^{(1)}_2 = 3e \mid \mathbf{E} \mid a_0. \tag{8.147}$$

Substituting these in the matrix equation (8.138), we determine,

$$S_{11} = S_{21} = \frac{1}{\sqrt{2}},$$

$$S_{12} = -S_{22} = \frac{1}{\sqrt{2}},$$

so that,

$$\Phi_1 = \frac{1}{\sqrt{2}} (u_{200} + u_{210}),$$

$$\Phi_2 = \frac{1}{\sqrt{2}} (u_{200} - u_{210}). \tag{8.148}$$

Substituting in Eqs. (8.144a, b) from Eqs. (8.146-148), we have,

$$E_1' = E^{(0)} - 3e \mid \mathbf{E} \mid a_0,$$

$$E_2' = E^{(0)} + 3e \mid \mathbf{E} \mid a_0,$$

$$E_{3,4}' = E^{(0)},$$

and

$$\psi_n' = \Phi_n, \ (n = 1, \ldots 4).$$

We see that the splitting is proportional to the strength of the applied electric field (hence the name *linear* Stark Effect). A diagrammatic representation of the effect is given in Fig. 8.7.

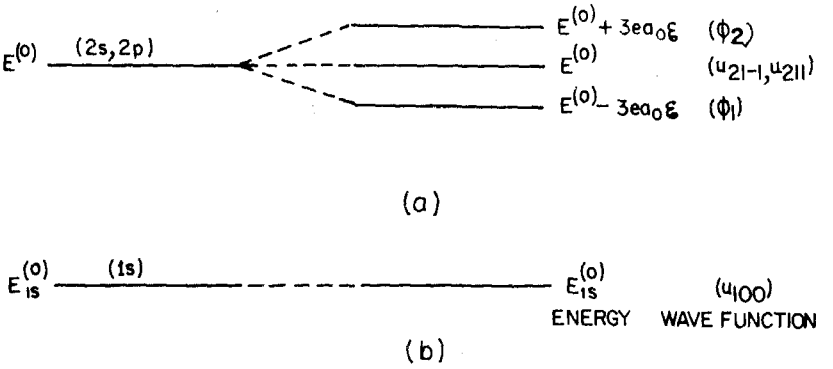


Fig. 8.7. Linear Stark Effect in hydrogen.
(a) Levels in the absence of the field,
(b) levels in the presence of the field.

Zeeman Effect in Hydrogen

The splitting of atomic levels, when the atom is placed in a uniform magnetic field, is known as Zeeman Effect. The effect of a magnetic field is to change the momentum from $\mathbf{p}$ to $(\mathbf{p} - \frac{e}{c}\mathbf{A})$, where $\mathbf{A}$ is the vector potential related to the field $\mathbf{B}$ by $\mathbf{B} = \text{curl } \mathbf{A}$, or, since the field is uniform ($\partial B/\partial x = 0$, etc.) and div. $\mathbf{B} = 0$, by[20] $\mathbf{A} = \frac{1}{2}(\mathbf{B} \times \mathbf{r})$. In the case of a weak field, the Hamiltonian of the hydrogen atom (neglecting spin) is, therefore, given by

$$\hat{H} = \hat{H}_0 + \hat{V},$$

with

$$\begin{aligned} \hat{V} &= -(e/2\mu c)(\hat{\mathbf{p}} \cdot \mathbf{A} + \mathbf{A} \cdot \hat{\mathbf{p}}) \\ &= -(e/\mu c)(\mathbf{A} \cdot \hat{\mathbf{p}}), \text{ since } \hat{\mathbf{p}} \cdot \mathbf{A} = \mathbf{A} \cdot \hat{\mathbf{p}} - i\hbar(\nabla \cdot \mathbf{A}) \\ &= -(e/2\mu c)(\mathbf{B} \times \hat{\mathbf{r}}) \cdot \hat{\mathbf{p}} = -(\mu_B/\hbar)(\mathbf{B} \cdot \hat{\mathbf{l}}), \end{aligned} \tag{8.149}$$

---

20. The identity,

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = \mathbf{A}(\nabla \cdot \mathbf{B}) - \mathbf{B}(\nabla \cdot \mathbf{A}) + (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B},$$

is used.

where,

$$\hat{\mathbf{l}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}},$$

is the orbital angular momentum, μ is the mass of the electron and

$$\mu_B = \frac{e\hbar}{2\mu c},$$

is the Bohr magneton.

By choosing **B** along the z-axis, we could make the perturbation diagonal with respect to the eigenvectors of $\hat{H}_0$. That is,

$$< u_{n'l'm'} \mid \hat{V} \mid u_{nlm} = -(\mu_B/\hbar)B < u_{n'l'm'} \mid \hat{l}_z \mid u_{nlm} >$$

$$= -\mu_B m B \delta_{nn'} \delta_{ll'} \delta_{mm'}.$$

Thus,

$$\Phi_k \equiv \Phi_{nlm} = \mathbf{u}_k;$$

and

$$E_k^{(1)} = -\mu_B B m. \tag{8.150}$$

Fig. 8.8 shows the splitting of the $1s$, $2s$, and $2p$ levels. Just as in the case of the Stark Effect, the degeneracy of the $(2s, 2p)$ level is only partially lifted. But we see from Figs. 8.7 and 8.8 that the levels degenerate in the Stark Effect are split in the Zeeman Effect, and vice versa.

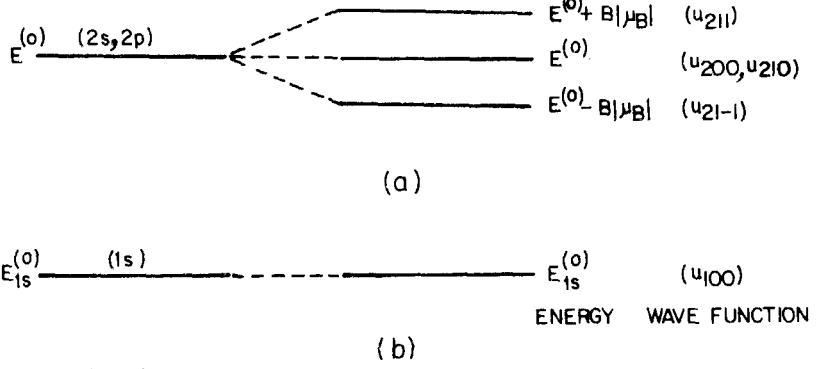


Fig. 8.8. Zeeman effect in hydrogen atom (neglecting spin of the electron).

Problem 8.10: When spin of the electron is taken into account, in place of Eq. (8.149), we have,

$$\hat{V} = -(\mu_B/\hbar)\mathbf{B} \cdot (\hat{\mathbf{l}} + \hat{\mathbf{s}})$$

Show that, in this case, a level of given total angular momentum quantum number j is split into $(2j+1)$ levels according to the formula (cf. Eq. (8.150)),

$$E_{jm}^{(1)} = -g_j \mu_B B m,$$

where the *Lande' factor* g_j is given by (here, $\mid nljm > \equiv \Phi_{nljm}$)

$$g_j = < nlj \mid\mid \hat{\mathbf{j}} \cdot (\hat{\mathbf{l}} + 2\hat{\mathbf{s}}) \mid\mid nlj > / \{j(j+1)\hbar^2\}.$$

Spin-Orbit Coupling in the Hydrogen Atom

When spin of the electron is taken into account, the degeneracy associated with a level of given orbital angular momentum l increases to $(2s+1)(2l+1) = 2(2l+1)$. This degeneracy is partly lifted by the interaction, or coupling, between the spin and the orbital motions of the electron. The spin-orbit interaction can be represented by a potential of the form, $\hat{V} = \alpha(\hat{\mathbf{l}}.\hat{\mathbf{s}})$ with α a constant for the atom, and can be treated as a perturbation since $|\hat{V}| \ll |\hat{H}_0|$. The zero-order wavefunctions are now products of a space part (u_{nlm_l}) and a spin part (x_{sm_s}). We will denote these by $|nlsm_lm_s>$. That is,

$$|u_k> \equiv \psi_k^{(0)} \equiv |nlsm_lm_s> = u_{nlm_l}x_{sm_s}. \tag{8.151}$$

The operator $(\hat{\mathbf{l}}.\hat{\mathbf{s}})$ does not commute with $\hat{l}_z$ and $\hat{s}_z$ (see Section (5.5A)). $\hat{V}$ is, thus, non-diagonal in the representation (the *uncoupled* representation) defined by the basis vectors (8.151). It is, therefore, necessary to construct the matrix of $\hat{V}$ in the subspace of the degenerate eigenfunctions and diagonalize it to obtain the Φ_k's. However, we can avoid this calculation by noting that $(\hat{\mathbf{l}}.\hat{\mathbf{s}}) = \frac{1}{2}(\hat{\mathbf{j}}^2 - \hat{\mathbf{l}}^2 - \hat{\mathbf{s}}^2)$, is diagonal in the *coupled* representation defined by the basis vectors (Eqs. (5.66a) and (5.67b)),

$$\Phi_q \equiv |nlsjm> = \sum_{m_l,m_s} C^{lsj}_{m_lm_sm} u_{nlm_l}x_{sm_s}, \tag{8.152}$$

where

$$\hat{\mathbf{j}} = \hat{\mathbf{l}} + \hat{\mathbf{s}},$$

and is the operator corresponding to the total angular momentum of the atom. We have, since $s = 1/2$,

$$<\Phi_{q'}|\hat{V}|\Phi_q> = \frac{\alpha\hbar^2}{2}[j(j+1) - l(l+1) - s(s+1)]\delta_{qq'}, \tag{8.153}$$

$$= -\left(\frac{l+1}{2}\right)\alpha\hbar^2, \quad j = l - \frac{1}{2} \tag{8.153a}$$

$$= \left(\frac{l}{2}\right)\alpha\hbar^2, \qquad j = l + \frac{1}{2}, \tag{8.153b}$$

where

$$q \equiv (nlsjm).$$

Thus, the $2(2l+1)$ degenerate levels belonging to a particular l-value will split into two levels (except when $l = 0$), one of $j = l - \frac{1}{2}$ and degeneracy $2\left(l - \frac{1}{2}\right) + 1 = 2l$, and the other of $j = l + \frac{1}{2}$ and degeneracy equal to $2(l+1)$. Fig. 8.9 shows the spin-orbit splitting of the level $(2s, 2p)$ in the Hydrogen atom, where

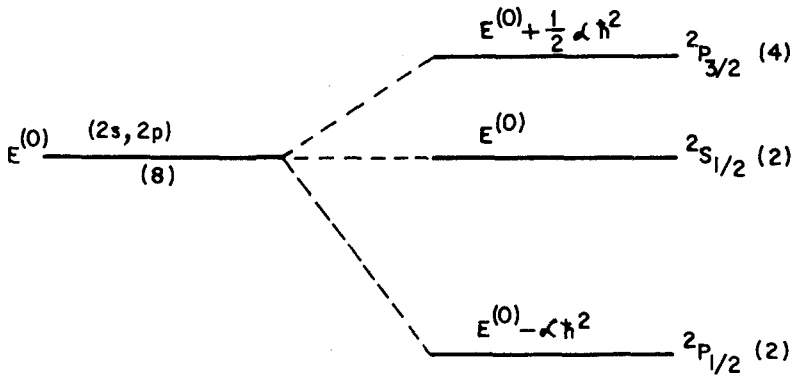


Fig. 8.9 Spin-orbit splitting of the $(2s, 2p)$ level in the hydrogen atom. The numbers in the brackets denote the degeneracy of the levels.

the notation $^{2S+1}L_J$ is used for labelling the levels (S = the spin, L = the orbital angular momentum (see Eq. (4.124) for the alphabetic notation for the l-value), and J = the total angular momentum).

Problem 8.11: Determine the C-coefficients in (8.152) for the case of the $2p$-level by diagonalizing the matrix of $\hat{V}$ in the uncoupled representation.

B. METHODS FOR TIME-DEPENDENT PROBLEMS

When the Hamiltonian is independent of time, the system remains in the state in which it finds itself at the beginning. In other words, there is no *transition* between different stationary states of the system. That is not the case when the Hamiltonian is time-dependent. If the system is in a state $|u_i\rangle$ at the initial time t_0, then it need not be found in this state at a later time t. The problem is then to determine the state vector (representing the state of the system) at time t, or, equivalently, to evaluate the probability for the system to be found in the state $|u_j\rangle$ at time t when it is known to have been in the state $|u_i\rangle$ at time t_0. An exact solution of the problem is rarely practical. Hence the necessity for resorting to approximation methods.

Now, time-dependence of the Hamiltonian can arise in one of the following two ways:

(1) $\hat{H} = \hat{H}_0 + \hat{V}(t)$, where $\hat{H}_0$ is independent of time, and

$$|\hat{V}(t)| \ll |\hat{H}_0|.$$

(2) $\hat{H}$ is constant in time except for a time T during which $\hat{H}$ changes from $\hat{H}_0$ to $\hat{H}_1$.

The approximation method which deals with the first case is known as the *time-dependent perturbation theory*. This is discussed in section 8.4. The other case is discussed in Section 8.5.

8.4. TIME-DEPENDENT PERTURBATION THEORY

Just as in the case of stationary perturbations (Eq. (8.108)), the Hamiltonian of the system can be separated into a major part $\hat{H}_0$ and a minor part $\hat{V}$. The only difference is that now $\hat{V}$ would be time-dependent, Thus,

$$\hat{H}(t) = \hat{H}_0 + \hat{V}(t). \tag{8.154}$$

$\hat{V}(t)$ will represent an *external* field (that is, one imposed from outside the system whose Hamiltonian is $\hat{H}_0$). For example, it could be an applied electromagnetic field, or it could be the interaction of a particle passing by the system, as in a scattering problem.

Instead of proceeding along the lines of Section 8.3[21], we adopt here a procedure which makes use of the evolution operator $\hat{U}(t, t_0)$ introduced in Section 4.1 (Eq. (4.1)). The problem of determining the state vector of the system at any time t, then, reduces to evaluating $\hat{U}(t, t_0)$ for all values of t.

Now, according to Problem 4.4,

$$\hat{U}(t, t_0) = \hat{U}_0(t, t_0)\hat{U}_I(t, t_0), \tag{8.155^1}$$

where,

$$\hat{U}_0(t, t_0) = \exp\,[-(i/\hbar)\hat{H}_0(t - t_0)], \tag{8.156}$$

and $\hat{U}_I(t, t_0)$ is defined by the equation,

$$i\hbar\frac{\partial \hat{U}_I(t, t_0)}{\partial t} = \hat{V}_I(t)\hat{U}_I(t, t_0), \tag{8.157}$$

with (see Eq. (4.40a)),

$$\hat{V}_I(t) = \hat{U}_0^{-1}(t, t_0)\hat{V}(t)\hat{U}_0(t, t_0). \tag{8.158}$$

Eq. (8.157) is equivalent to the integral equation,

$$\hat{U}_I(t, t_0) = \hat{1} - (i/\hbar)\int_{t_0}^{t} \hat{V}_I(t_1)\hat{U}_I(t_1, t_0)dt_1. \tag{8.159^1}$$

Eq. (8.159^1) enables us to express $\hat{U}_I$ as a power series in the perturbation $\hat{V}_I$ using the method of iteration (See Section 7.4, Eqs. (7.107$^{1\text{-}2}$)). The first iterated form of $\hat{U}_I$ is obtained by substituting for $\hat{U}_I(t_1, t_0)$ in (8.159^1) the expression,

$$\hat{U}_I(t_1, t_0) = \hat{1} - (i/\hbar)\int_{t_0}^{t_1} \hat{V}_I(t_2)\hat{U}_I(t_2, t_0)dt_2\,,$$

which is obtained by substituting t_1 for t in that equation. Thus

$$\hat{U}_I(t, t_0) = \hat{1} + (i\hbar)^{-1}\int_{t_0}^{t} \hat{V}_I(t_1)dt_1 + (i\hbar)^{-2}\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \cdot \hat{V}_I(t_1)\hat{V}_I(t_2)\hat{U}_I(t_2, t_0). \tag{8.159^2}$$

The second iterated form results when $\hat{U}_I(t_2, t_0)$ in this equation is replaced by the expression obtained by replacing t by t_2 in Eq. (8.159^1). Repetition of this procedure leads to the result,

---

21. For such a procedure see Schiff, L.I., *Quantum Mechanics*, III Edition, Section 35.

$$\hat{U}_I(t,t_0) = \sum_{n=0}^{\infty} \hat{U}_I^{(n)}(t,t_0), \tag{8.159^3}$$

with,

$$\hat{U}_I^{(n)}(t,t_0) = (i\hbar)^{-n} \int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2 \ldots \int_{t_0}^{t_{n-1}} dt_n \cdot \hat{V}_I(t_1)\hat{V}_I(t_2)\ldots\hat{V}_I(t_n), \tag{8.160a}$$

where, $t_1 > t_2 > \ldots > t_{n-1} > t_n$,

and
$$\hat{U}_I^{(0)}(t,t_0) = \hat{1}. \tag{8.160b}$$

From Eqs. (8.155^1) and (8.159^3), we have,

$$\hat{U}(t,t_0) = \sum_{n=0}^{\infty} \hat{U}^{(n)}(t,t_0), \tag{8.155^2}$$

where, $\hat{U}^{(n)}(t,t_0) = \hat{U}_0(t,t_0)\hat{U}_I^{(n)}(t,t_0)$

$$= (i\hbar)^{-n} \int_{t_0}^{t} dt_1 \ldots \int_{t_0}^{t_{n-1}} dt_n \hat{U}_0(t,t_0).\hat{V}_I(t_1)\ldots\hat{V}_I(t_n)$$

$$= (i\hbar)^{-n} \int_{t_0}^{t} dt_1 \ldots \int_{t_0}^{t_{n-1}} dt_n \hat{U}_0(t,t_1)\hat{V}(t_1)$$

$$.\hat{U}_0(t_1,t_2)\hat{V}(t_2)\ldots\hat{U}_0(t_{n-1},t_n)\hat{V}(t_n)\hat{U}_0(t_n,t_0). \tag{8.161}$$

Here, properties (4.9) and (4.7) of the evolution operator and definition (8.158) of $\hat{V}_I$ have been used. Eq. (8.155^2) is, thus, an expansion in powers of the perturbation $\hat{V}(t)$. Correspondingly, the wave function of the system at time t is given by the series,

$$\Psi(t) \equiv \hat{U}(t,t_0)\Psi_0 = \sum_{n=0}^{\infty} \Psi^{(n)}(t), \tag{8.162}$$

where,
$$\Psi_0 \equiv \Psi(t_0)$$

and,
$$\Psi^{(n)}(t) = \hat{U}^{(n)}(t,t_0)\Psi_0. \tag{8.163}$$

Here, $\Psi^{(0)}(t) = \hat{U}_0(t,t_0)\Psi_0$, is the zero-order wave-function and $\Psi^{(n)}(t), (n \geq 1)$ the nth *order* correction to the wavefunction. From Eqs. (8.161) and (8.163), we see that $\Psi^{(n)}(t)$ involves n change of states, in its evolution from t_0 to t. This is illustrated in Fig. 8.10 where the Feynman graphs[22] corresponding to the integrands of $\Psi^{(n)}(t)$ for $n = 0$, 1 and 2 are shown. Suppose Ψ_0 is an eigenvector of $\hat{H}_0$. Then, in the case of $\Psi^{(2)}$, the system evolves as Ψ_0 between t_0 and t_1; at t_1 it makes, under the influence of $\hat{V}(t_1)$ a transition to a new eigenstate of $\hat{H}_0$. The system remains in this new state until t_2 when it again makes a transition to another

22. A description of Feynman Graphs, or Feynman Diagrams, is given in Section 11.5.

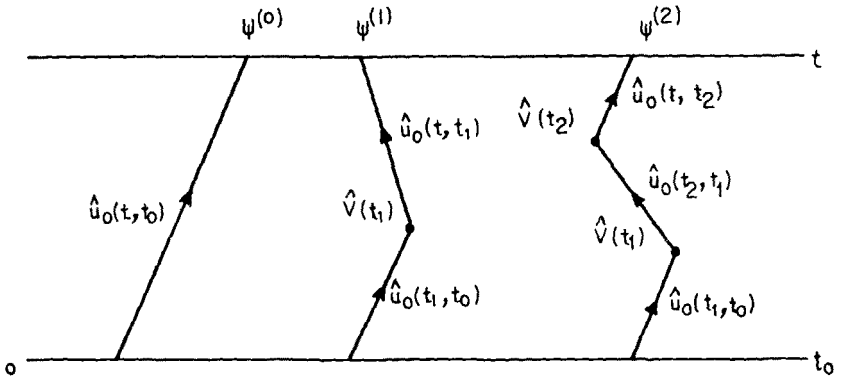


Fig. 8.10. Feynman graphs of the zero, first and second order contributions to $\psi(t)$. The evolution operators appropriate to each segment of the graphs are also shown.

eigenstate of $\hat{H}_0$ under the influence of $\hat{V}(t_2)$. The evolution of the system between t_0 and t_1, t_1 and t_2, and t_2 and t are, thus, described by the evolution operators appropriate to $\hat{H}_0$ while the change of states at t_1 and t_2 are determined by the perturbation $\hat{V}$. When $|\hat{V}(t)|$ is small enough to be treated as a perturbation, the amplitude of the state would decrease rapidly with each change of state. In the case of an ensemble, this means that the number of systems (members of the ensemble[23]) that would undergo a change of state would be only a small fraction of the total number present at that time. Thus, $\Psi^{(0)}$ represents systems that have evolved without making any change of state, $\Psi^{(1)}$ systems that have changed state once, $\Psi^{(2)}$ those that have changed states twice, and so on. It follows that the series (8.162) would be a fast-converging one. The approximation to order n consists in neglecting contributions to $\Psi(t)$ from systems that have undergone change of state more than n times between t_0 and t. That is, in setting,

$$\Psi(t) \approx \sum_{p=0}^{n} \Psi^{(p)}(t) = \sum_{p=0}^{n} \hat{U}^{(p)}(t, t_0)\Psi_0. \qquad (8.162')$$

In the following, we will confine the discussion to those cases where the *first order* approximation is good enough. Also, we will assume that the system is initially in an eigenstate of $\hat{H}_0$. Then, the problem is to determine the probability for the system to be in an eigenstate of $\hat{H}_0$ at time t, that is different from the initial state.

Let E_k and ψ_k represent, respectively, the eigenvalues (assumed to be discrete) and (the normalized) eigenvectors of $\hat{H}_0$:

$$\hat{H}_0\psi_k = E_k\psi_k, \quad k = 0, 1, 2, \ldots, \qquad (8.164$$

with (see Eq. (4.21)),

$$\psi_k(\mathbf{r}, t) = u_k(\mathbf{r}) \exp[-(i/\hbar)E_k(t - t_0)]. \qquad (8.164a$$

Also, let the system be in the state ψ_i at time t_0. Then, in the first-order approximation the probability for the system to be found in the state ψ_f at time t, is given by

23. These are, for example, the individual atoms in an ensemble of atoms.

$$w_{i\to f} = |< \psi_f(t) \mid \Psi(t) >|^2, \tag{8.165^1}$$

where, from Eq. (8.162^1),

$$\langle \psi_f \mid \Psi(t) \rangle \approx \langle \psi_f \mid \hat{U}^{(0)}(t, t_0) \mid \psi_i \rangle + \langle \psi_f \mid \hat{U}^{(1)}(t, t_0) \mid \psi_i \rangle \tag{8.166}$$

Now,

$$\langle \psi_f \mid \hat{U}^{(0)}(t, t_0) \mid \psi_i \rangle = \exp\,[+(i/\hbar)(E_f - E_i)(t - t_0)]\, \langle u_f \mid u_i \rangle$$

$$= \delta_{if}, \tag{8.166a}$$

(by Eqs. (8.161), (8.160b), (8.156) and (6.164a)),

and

$$\langle \psi_f \mid \hat{U}^{(1)}(t, t_0) \mid \psi_i \rangle = \frac{1}{i\hbar} \int_{t_0}^{t} dt_1 \langle \psi_f \mid \hat{U}_0(t, t_1) \hat{V}(t_1) \hat{U}_0(t_1, t_0) \mid \psi_i \rangle$$

$$= \frac{1}{i\hbar} \int_{t_0}^{t} V_{fi}(t_1) \exp\,[i\omega_{fi}(t_1 - t_0)]\, dt_1, \tag{8.166b}$$

where,

$$V_{fi}(t_1) = \langle u_f \mid \hat{V}(t_1) \mid u_i \rangle, \tag{8.167a}$$

and

$$\omega_{fi} = (E_f - E_i)/\hbar, \tag{8.167b}$$

is the *Bohr frequency* corresponding to the transition $\psi_i \to \psi_f$. Substituting from Eqs. (8.166) and (8.166a, b) in (8.165^1), we get,

$$w_{i\to f} = \frac{1}{\hbar^2} \left| \int_{t_0}^{t} V_{fi}(t') \exp\,(i\omega_{fi} t') dt' \right|^2, (i \neq f). \tag{8.165^2}$$

This is the basic equation of the *first-order time-dependent perturbation theory.* We will consider certain illustrative applications of this formula in three different types of time-dependence.

Problem 8.12: Show that Eq. (8.160a) can also be written as

$$\hat{U}_I^{(n)}(t, t_0) = \frac{1}{(i\hbar)^n n!} \hat{P} \int_{t_0}^{t} dt_1 \int_{t_0}^{t} dt_2 \ldots \int_{t_0}^{t} dt_n \hat{V}_I(t_1).\hat{V}_I(t_2) \ldots \hat{V}_I(t_n),$$

where $\hat{P}$ is defined by,

$$\hat{P}\{\hat{A}(t_1)\,\hat{B}(t_2)\} = \begin{cases} \hat{A}(t_1)\hat{B}(t_2), & \text{if } t_1 > t_2, \\ \hat{B}(t_2)\hat{A}(t_1), & \text{if } t_2 > t_1. \end{cases}$$

Note: $\hat{P}$ is known as the *Dyson chronological operator* (See Section 11.5).

8.4A. Constant Perturbation

This is the case when $\hat{V}$ has no explicit dependence on time. Then, from Eq. (8.165^2), we have (setting $t_0 = 0$),

$$w_{i\to f} = \{|\, V_{fi}\,|^2/\hbar^2\} \cdot \left| \int_0^t [\exp\,(i\omega_{fi} t')]\, dt' \right|^2$$

$$= \frac{2|V_{fi}|^2}{\hbar^2\omega_{fi}^2} \cdot (1 - \cos\omega_{fi}t)$$

$$= \frac{4|V_{fi}|^2}{\hbar^2\omega_{fi}^2} \cdot \sin^2\left(\frac{\omega_{fi}}{2}t\right). \tag{8.168}$$

In Fig. 8.11(a) is plotted the variation of $w_{i\to f}$ for a given (non-zero) value of ω_{fi} as a function of t, while in Fig. 8.11(b) we show the variation of $w_{i\to f}(t)$ with ω_{fi}.

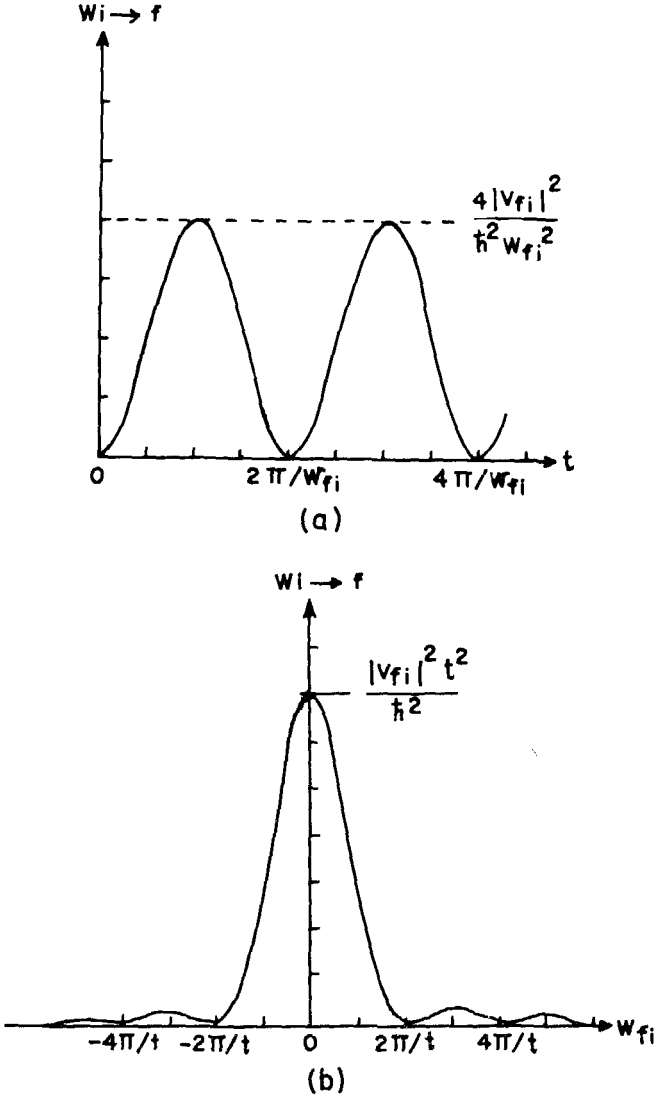


Fig. 8.11 Variation of $w_{i\to f}$ [Eq. (8.168)] with; (a) t, (b) ω_{fi}. The secondary maxima in (b) are equal to $4V_{fi}^2/\hbar\omega_{fi}^2$.

We see that the probability oscillates in time (except for $\omega_{fi} = 0$) between zero and a maximum of $4 | V_{fi} |^2 / (E_f - E_i)^2$ with a period τ characteristic of the energy change $E_f - E_i$ (that is, $\tau = 2\pi/\omega_{fi}$). Since $| V_{fi} |$ is small, even this maximum value would be negligible except when ω_{fi} is small compared with $| V_{fi} |$. In fact, we see from Fig. 8.11 (b) that the probability $w_{i \to f}$ is appreciable only for a range of values of ω_{fi} that lies within a band of width,

$$\Delta\omega_{fi} \approx 2\pi/t, \tag{8.169a}$$

around the value $\omega_{fi} = 0$. That is, transitions take place mostly to those states that have their energy within $\Delta E \sim h/t$ of the energy of the initial state. The condition for the perturbation to induce a transition is, thus,

$$\Delta E \cdot t < h, \text{ or } t < \tau, \tag{8.169b}$$

where ΔE is the change in the energy of the system, t the duration of perturbation and τ the characteristic time associated with the energy change ΔE.

Result (8.169b) is some times interpreted, on the basis of a time-energy uncertainty relationship, as signifying that a constant perturbation rarely causes transitions between states having *measurably different* energies. The existence of a time-energy uncertainty relationship is not, however, beyond dispute (See Section 3.2).

Transition to a Continuum

Since the transition probability is appreciable only for states having energies nearly equal to the energy of the initial state, the above procedure is best suited when the energy levels (to which transitions take place) are part of a continuum. We could still think of the levels to be discrete but infinitesimally close together. It is then possible to define a transition probability per unit time.

Let $\rho(E_f)$ be the *density of states* (the number of energy levels per unit energy interval) in the neighbourhood of (and including) the level $| u_f >$. Then, $\rho(E_f)dE_f$ is the number of states having energies between E_f and $E_f + dE_f$. The product, $w_{i \to f}\rho(E_f)dE_f \equiv dW$, is the probability for transition to this group of levels; and the *total* transition probability for transition to the continuum states is given by,

$$W = \int dW = \int w_{i \to f}\rho(E_f)dE_f$$

$$= 4 \int_{-\infty}^{+\infty} | V_{fi} |^2 \rho(E_j) \frac{\sin^2(\omega_{fi}/2)t}{\hbar\omega_{fi}^2} d\omega_{fi}, \tag{8.170^1}$$

where Eq. (8.168) has been used. Now, both $\rho(E_f)$ and V_{fi} are slowly varying functions of the energy. And since the contribution to the integral in (8.170^1) comes (according to Fig. 8.11(b)) from a narrow band of energy, the factor $| V_{fi} |^2 \rho(E_f)$ could be taken outside the integral sign. Then,

$$W = 4 \mid V_{fi} \mid^2 \rho(E_f)\,(t/2\hbar) \int_{-\infty}^{+\infty} x^{-2} \sin^2 x \, dx, \left(x = \frac{1}{2}\omega_{fi} t\right),$$

$$= \left(\frac{2\pi}{\hbar}\right) \rho(E_f) \mid V_{fi} \mid^2 t. \qquad (8.170^2)$$

The transition probability per unit time is, then,

$$\frac{dW}{dt} = \frac{2\pi}{\hbar} \mid V_{fi} \mid^2 \rho(E_f). \qquad (8.171^1)$$

Thus, the rate of transition is independent of time. This result is known as the *Golden Rule of time-dependent perturbation theory.* The proportionality of W to t could have been inferred from Fig. 8.11(b). Since the height of the main peak is proportional to t^2 and width inversely proportional to t, the area under the curve is proportional to t.

Eventhough expression (8.171^1) is derived as the sum of the transition probabilities per unit time to all the levels in the continuum, by virtue of the assumptions involved in going from (8.170^1) to (8.170^2), it could be interpreted as the transition probability per unit time for the transition $\mid u_i\rangle \rightarrow \mid u_f\rangle$. The dependence of (8.171^1) on $\mid u_f\rangle$ comes through the matrix element $\mid V_{fi} \mid$. Now, the condition for the validity of expression (8.170^2) is that the contribution to the integral in (8.170^1) come from a very narrow band of energy including the energy E_i (so that the range $-\infty$ to $+\infty$ is, in effect, the range $(E_i - \epsilon)$ to $(E_i + \epsilon)$, where, ϵ is infinitesimal). We see, from Fig. 8.11(b), that this requires t to be large and that, in that case, the energy is practically *conserved* in the transition. The condition for the validity of Eq. (8.171^1) can be, thus, incorporated into the equation by multiplying it by the factor $\delta(E_i - E_f)$. Eq. (8.171^1) then reads:

$$\frac{dW}{dt} \equiv (W_c)_{i \rightarrow f} = \frac{2\pi}{\hbar} \mid V_{fi} \mid^2 \rho(E_f)\delta(E_i - E_f). \qquad (8.171^2)$$

Here, the subscript c emphasizes the fact that the states $\mid u_i\rangle$ and $\mid u_f\rangle$ are parts of a continuum, unlike the states involved in Eq. (8.168).

Scattering Cross Section in the Born Approximation

As an example of the application of formula (8.171^2), we will calculate the scattering cross-section in the (first) Born Approximation (Eq. (7.106b)). Now, under conditions of validity of the Born approximation, the kinetic energy of the particles (being scattered) are very much larger than the scattering potential. The particles are free long before, and long after, the scattering, and come under the influence of the potential for a certain duration in between. From the viewpoint of the particles, therefore, the potential could be treated as a time dependent, but constant, perturbation. The unperturbed Hamiltonian, in this case, is the free-particle Hamiltonian:

$$\hat{H}_0 = \frac{\hat{p}^2}{2\mu} = -\frac{\hbar^2}{2\mu}\nabla^2.$$

The eigenvectors of $\hat{H}_0$ are the 'plane waves' $e^{i\mathbf{k}\cdot\mathbf{r}}$, where $\mathbf{p}=\hbar\mathbf{k}$, whereas the eigenvalues are given by $E(k)=\hbar^2k^2/2\mu$. Since k can vary continuously between 0 and ∞, the spectrum of $\hat{H}_0$ forms a continuum.

Also, in elastic scattering, the energy of the particle is the same both before and after scattering so that energy is conserved in the transition from the initial to the final state. This is, therefore, a suitable case for the application of formula (8.171[2]).

We can make the eigenvalues discrete by enclosing the particle in a cubic box of volume L^3 and imposing periodic boundary conditions[24]. Then, the normalized eigenvectors of $\hat{H}_0$, representing the initial and final states are given by,

$$u_i(\mathbf{r}) \equiv <\mathbf{r} \mid u_i> = L^{-3/2}e^{i\mathbf{k}_i\cdot\mathbf{r}}, \tag{8.172a}$$

$$u_f(\mathbf{r}) \equiv <\mathbf{r} \mid u_f> = L^{-3/2}e^{i\mathbf{k}_f\cdot\mathbf{r}}, \tag{8.172b}$$

so that,

$$\begin{aligned} V_{fi} &= \int u_f^*(\mathbf{r})V(\mathbf{r})u_i(\mathbf{r})d^3\mathbf{r} \\ &= \frac{1}{L^3}\int V(\mathbf{r})e^{i\mathbf{K}\cdot\mathbf{r}}d^3\mathbf{r}, \end{aligned} \tag{8.173}$$

where,

$$\hbar\mathbf{K} = \hbar(\mathbf{k}_i - \mathbf{k}_f),$$

is the momentum transfer.

The density of states $\rho(E_f)$ could be found as follows:

Let the time-dependent Schrödinger equation corresponding to $\hat{H}_0$ be

$$i\hbar\frac{\partial\psi(\mathbf{r},t)}{\partial t} = \hat{H}_0\psi(\mathbf{r},t) = (-\hbar^2/2\mu)\nabla^2\psi.$$

Then (see Eq. (4.21)), $\psi(\mathbf{r},t) = u(\mathbf{r})\exp[-(i/\hbar)Et]$. The normalization of ψ should be independent of time. That is,

$$\begin{aligned} 0 = \frac{\partial}{\partial t}\int\psi^*\psi d^3\mathbf{r} &= \int\left(\psi^*\frac{\partial\psi}{\partial t} + \frac{\partial\psi^*}{\partial t}\psi\right)d^3\mathbf{r} \\ &= \frac{i\hbar}{2\mu}\int[\psi^*(\nabla^2\psi) - (\nabla^2\psi^*)\psi]d^3\mathbf{r} \\ &= -\int(\nabla\cdot\mathbf{j})d^3\mathbf{r} = -\oint_A j_n d^3\mathbf{r}, \end{aligned} \tag{8.174}$$

where,

$$\mathbf{j} = \frac{\hbar}{2i\mu}[\psi^*(\nabla\psi) - (\nabla\psi^*)\psi], \tag{8.175}$$

---

24. Such a procedure leads to a restriction in the allowed values of k. But the spacing between adjacent values of k could be made as small as desired by increasing L [see Eq. (8.177) below].

and j_n is the component of $\mathbf{j}$ along the outward normal to the closed surface A enclosing the volume over which the integral is taken ($d^2\mathbf{r}$ denotes an element of this surface). The last line in (8.174) follows from the second on partial integration while the last step results from the application of the *divergence theorem*[25].

Now, Eq. (8.174) will be satisfied if either ψ vanishes everywhere on the bounding surface or ψ is periodic (in a certain way) on this surface. In the present case, $u(\mathbf{r})$ defined by Eq. (8.172) does not vanish anywhere. So condition (8.174) should be met by periodic boundary conditions. In the case of the cubic volume, these boundary conditions are that $u(\mathbf{r})$ and the normal component of its derivatives should be the same at opposite faces of the cube. Choosing the edges of the cube along the co-ordinate axes, we have,

$$\exp(ik_x x) = \exp[ik_x(x+L)],$$

or,

$$k_x L = 2\pi n_x, \text{ with } n_x = 0, \pm 1, \pm 2, \ldots. \tag{8.176a}$$

Similarly,

$$k_y L = 2\pi n_y, \tag{8.176b}$$

$$k_z L = 2\pi n_z, \tag{8.176c}$$

and

$$kL = 2\pi n = 2\pi(n_x^2 + n_y^2 + n_z^2)^{1/2}, \tag{8.177}$$

where, n_y and n_z also are positive or negative integers, including zero.

We have to find the number of states between $E(k)$ and $E(k+dk)$, where the allowed values of k are given by Eq. (8.177). This is equal to the number of points in the n-space between n and $(n+dn)$ with co-ordinates (n_x, n_y, n_z) integer. Now, in the n-space there is one such point per unit volume. Therefore, the number of k-values between k and $(k+dk)$ in the k-space is

$$4\pi n^2 dn = 4\pi(L/2\pi)^3 k^2 dk.$$

This number refers to particles scattered in all directions; that is, over a solid angle 4π. Therefore, the number of states with k-values in the interval k_f and $k_f + dk_f$ within the solid angle[26] $d\Omega$ is given by

$$(L/2\pi)^3 k_f^2 dk_f d\Omega \equiv \rho(E_f) dE_f = \rho(E_f) \cdot (\hbar^2 k_f/\mu) dk_f;$$

so that

$$\rho(E_f) = \frac{\mu}{\hbar^2}\left(\frac{L}{2\pi}\right)^3 k_f d\Omega. \tag{8.178}$$

---

25. In view of the interpretation of $\psi^*\psi$ as the probability density, $\mathbf{j}$ is interpreted as the *probability current density*. Eq. (8.174), then, represents the conservation of probability or, equivalently, the conservation of number of particles.

26. Unless we restrict the states to a small solid angle, they will not all be described by the *same* wavefunction which, in this case, is characterised by the momentum vector $\mathbf{k}_f$.

Substituting from (8.173) and (8.178) in (8.171[2]), we get,

$$(W_c)_{i\to f}=\frac{2\pi\mu k_f}{(2\pi\hbar L)^3}\left|\int V(\mathbf{r})\exp(i\mathbf{K}\cdot\mathbf{r})d^3\mathbf{r}\right|^2 d\Omega.$$

This is equal to the number dN of particles scattered into $d\Omega$ per unit time when there is only one incident particle in L^3. Thus, incident flux $\mathbf{J}_0$ is given by

$$\mathbf{J}_0=\rho_i\mathbf{v}_i=\frac{1}{L^3}\cdot\left(\frac{\hbar\mathbf{k}_i}{\mu}\right).$$

The differential scattering cross-section $\sigma(\Omega)$ is, then, given by (see Eq. (7.1[1])), since $|\mathbf{k}_i|=|\mathbf{k}_f|=k$,

$$\sigma(\Omega)=\frac{dN}{|\mathbf{J}_0|\,d\Omega}=\left(\frac{\mu L^3}{\hbar k}\right)\frac{(W_c)_{i\to f}}{d\Omega}$$

$$=\left|\frac{1}{4\pi}\frac{2\mu}{\hbar^2}\int\exp(i\mathbf{K}\cdot\mathbf{r})V(\mathbf{r})d^3\mathbf{r}\right|^2, \tag{8.179}$$

which is in agreement with the Born approximation scattering amplitude (7.106b).

Problem 8.13: $\mathbf{u}_1$ and $\mathbf{u}_2$ are degenerate eigenvectors of the Hamiltonian $\hat{H}_0$ of a system. The introduction of a constant perturbation splits the two levels a distance ϵ in energy apart. If the system is initially in the state $\mathbf{u}_1$, obtain the condition for the system to be found in the state $\mathbf{u}_1$ even after the perturbation has been on for a time T.

8.4B. Harmonic Perturbation

A field which varies harmonically in time can be represented by

$$\hat{V}(t)=\hat{\mathrm{v}}(\mathbf{r})e^{i\omega t}+\hat{\mathrm{v}}^\dagger(\mathbf{r})e^{-i\omega t}, \tag{8.180}$$

where $\hat{\mathrm{v}}$ is independent of t (the second term in (8.180) is needed to make $\hat{V}$ Hermitian). Substituting in (8.165[2]) (with $t_0=0$), we get, in place of (8.168),

$$w_{i\to f}=\frac{4|\mathrm{v}_{fi}|^2}{\hbar^2}\left\{\sin\frac{[(\omega+\omega_{fi})/2]t}{\omega+\omega_{fi}}\right\}^2+\frac{4|\mathrm{v}_{if}|^2}{\hbar^2}\left\{\frac{\sin[(\omega-\omega_{fi})/2]t}{\omega-\omega_{fi}}\right\}^2+\ldots \tag{8.181}$$

Arguments similar to those in the case of Eq. (8.168) show that $w_{i\to f}$ is appreciable only when $\omega_{fi}\approx\pm\omega$. That is, when $|\omega_{fi}|\equiv|(E_j-E_i)|/\hbar=\omega$. But $|(E_f-E_i)|/\hbar$ is the *Bohr frequency* corresponding to the energy change $\Delta E=\pm(E_f-E_i)$. Thus, only those transitions are allowed for which the Bohr frequency is equal to the frequency of the perturbing field. $\omega_{fi}=\omega$ corresponds to *absorption* of energy from the field (transition from a state of lower energy to one of higher energy). In this case, the first term of the R.H.S. of (8.181) is negligible, so that, we have,

$$w_{i\to f}^{(a)} \approx \left(\frac{4\,|\,v_{if}\,|^2}{\hbar^2}\right)\frac{\sin^2\frac{\omega' t}{2}}{\omega'^2},\ (\omega' = \omega - \omega_{fi}). \tag{8.181a}$$

The case, $\omega_{fi} = -\omega$, on the other hand, corresponds to *emission*[27] of an amount of energy equal to $\hbar\omega$ by the system (transition from a state of higher energy to a lower one). The corresponding transition probability is given by,

$$w_{i\to f}^{(e)} \approx \left(\frac{4\,|\,v_{fi}\,|^2}{\hbar^2}\right)\cdot\frac{\sin^2\frac{\omega' t}{2}}{\omega'^2},\ (\omega' = \omega + \omega_{fi}) \tag{8.181b}$$

A plot of $w_{i\to f}$ against ω_{fi} is given in Fig. 8.12. The similarity of the peaks in Fig. 8.12 to the peak in Fig. 8.11(b) is obvious. Using arguments similar to those leading to Eq. (8.171[2]) we get, for the transition probability per unit time for transition from an initial state $|\,u_i>$ to a final state $|\,u_f>$ having energy approximately equal to $E_i + \hbar\omega$, the expression,

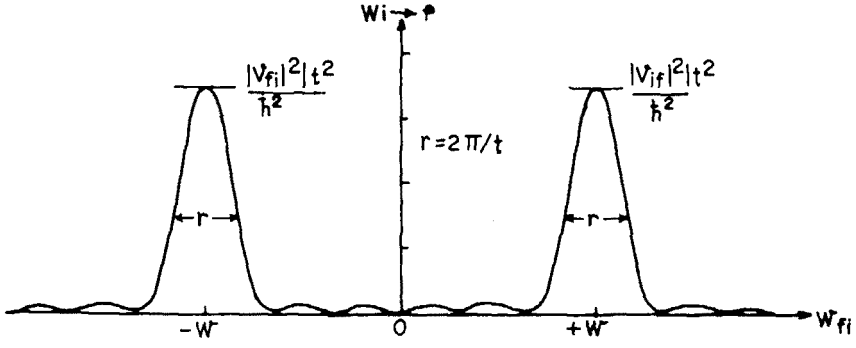


Fig. 8.12. Variation of $w_{i\to f}$ [Eqs. (8.181a) and (8.181b)] with ω_{fi}.

$$(w_c)_{i\to f}^{(a)} = \frac{2\pi}{\hbar}\,|\,v_{if}\,|^2\,\rho(E_f)\delta(E_j - E_i - \hbar\omega). \tag{8.182}$$

This formula is applicable when the initial state is part of a discrete spectrum and the final state is part of a continuum. Ionization of an atom in which an electron occupying the ground state of the atom absorbs a quantum of radiation from a perturbing electromagnetic field and jumps to the positive energy (continuum) part of the spectrum, is an example (Problem 8.14).

Problem 8.14: A hydrogen atom in its ground state, is subjected to an oscillating electric field, $\mathbf{E} = 2\mathbf{E}_0 \sin\omega t$, where ω is greater than the ionization frequency $[= \mu e^4/2\hbar^2$, (see Eq. 4.117)] of the atom. Obtain the differential ionization probability (that is, the probability per unit time per unit solid angle that the electron of the atom is ejected in the direction $\Omega \equiv (\theta, \phi)$), assuming that the final state of the electron can be represented by a plane wave.

---

27. In order to distinguish this process, whose driving force is an applied external field, from the process of *spontaneous emission* arising from fluctations in the internal fields of the system, the former is usually referred to as *induced, or stimulated, emission.*

When the transition is between discrete levels, however, the rate of transition is not constant in time, but is oscillatory (Fig. 8.11(a)), if the perturbing field is strictly monochromatic. The latter condition is rarely fulfilled in practice. More often than not, the field would consist of a range of frequencies. In such cases, it is possible to define a transition probability per unit time that is independent of time, as illustrated in the following example of atomic radiation.

Radiative Transitions in Atoms

Consider an atom (atomic number Z) placed in an external electromagnetic field. If the Hamiltonian of the atom in the absence of the field is

$$\hat{H}_0 = \sum_{j=1}^{z} \frac{1}{2m} \hat{\mathbf{p}}_j^2 + \hat{U}, \tag{8.183a}$$

then, the Hamiltonian in the presence of the field would be given by[28],

$$\hat{H} = \sum_{j=1}^{z} \left\{ \frac{1}{2m} \left(\hat{\mathbf{p}}_j - \frac{e}{c} \mathbf{A}(\mathbf{r}_j, t) \right)^2 + e\phi(\mathbf{r}_j, t) \right\} + \hat{U}. \tag{8.183b}$$

Here, $\hat{\mathbf{p}}_j$ represents the momentum of the jth electron, e and m the charge and the mass of an electron and $\mathbf{A}(\mathbf{r}_j)$ and $\phi(\mathbf{r}_j)$, respectively, the vector and the scalar potentials corresponding to the field at the position of the jth electron. Also, c is the velocity of light in vacuum. In terms of the potentials, the electric field $\mathbf{E}$ and the magnetic induction $\mathbf{B}$ are given by

$$\mathbf{E} = -\frac{1}{c}\frac{\partial A}{\partial t} - \nabla\phi; \quad \mathbf{B} = \nabla \times \mathbf{A}. \tag{8.184}$$

When the source of the field (charges and currents) are away from the atom, we can choose,

$$\phi = 0; \text{ div } \mathbf{A} = 0. \tag{8.185}$$

In this case, the vector potential satisfies the wave equation,

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0. \tag{8.186}$$

The plane-wave solution of Eq. (8.186) is given by,

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 \exp\{i(\mathbf{k}\cdot\mathbf{r} - \omega t)\} + \mathbf{A}_0^* \exp\{-i(\mathbf{k}\cdot\mathbf{r} - \omega t)\}, \tag{8.187}$$

where $\mathbf{A}_0$ is a constant vector perpendicular to the *propagation vector* $\mathbf{k}$ (the result, $\mathbf{k}\cdot\mathbf{A}_0 = 0$ is required to satisfy the *Lorentz condition* (8.185)), and

$$\omega = c\,|\,\mathbf{k}\,|\,. \tag{8.188}$$

Now from Eqs. (8.183a, b) and (8.185), we have (assuming $|\,\mathbf{A}\,|$ to be small),

$$\hat{V}(t) \equiv \hat{H} - \hat{H}_0 = -\sum_{j=1}^{z} \frac{e}{mc} (\mathbf{A}(\mathbf{r}_j, t)\cdot\hat{\mathbf{p}}_j), \tag{8.189}$$

where, the result (see Eq. (3.14d)),

$$\hat{\mathbf{p}}\cdot\mathbf{A} = \mathbf{A}\cdot\hat{\mathbf{p}} - i\hbar\nabla\cdot\mathbf{A},$$

---

28. See, Jackson, J.D., *Classical Electrodynamics* (John Wiley, New York (1963)) Section 12.5.

has been used. Substituting for **A** from Eq. (8.187), Eq. (8.189) reduces to the form (8.180) with

$$\hat{\mathrm{v}}(\mathbf{r}) = -\sum_{j=1}^{z} \frac{e}{mc}(\mathbf{A}_0^* \cdot \hat{\mathbf{p}}_j) \exp(-i\mathbf{k} \cdot \mathbf{r}_j). \tag{8.190}$$

Thus, the perturbation is proportional to the amplitude of the field. Therefore, the transition probability, which is proportional to the square of the matrix element of the perturbation, would be proportional to the intensity of the field. Now, the intensity is given by the Poynting vector:

$$\mathbf{S} = \frac{c}{4\pi}\mathbf{E} \times \mathbf{B}. \tag{8.191^1}$$

But, from eqs. (8.184) and (8.187), we have,

$$\mathbf{E} = -\frac{1}{c}\frac{\partial A}{\partial t} = i(\omega/c)\,[\mathbf{A}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)] - \mathbf{A}_0^* \exp[-i(\mathbf{k} \cdot \mathbf{r} - \omega t)]\,]$$

and $$\mathbf{B} = \nabla \times \mathbf{A} = \frac{\mathbf{k} \times \mathbf{E}}{k}, \text{ since } (\omega/c) = k,$$

so that,

$$\mathbf{S} = \frac{c}{4\pi k}\mathbf{E} \times (\mathbf{k} \times \mathbf{E}) = \frac{c}{4\pi} \cdot (\mathbf{k}/k)\mathbf{E}^2$$

$$= (\mathbf{k}/k) \cdot \frac{\omega^2}{4\pi c}[2 \mid \mathbf{A}_0 \mid^2 - \mathbf{A}_0^2 \exp\{2i(\mathbf{k} \cdot \mathbf{r} - \omega t)\}$$

$$-\mathbf{A}_0^{*2} \exp\{-2i(\mathbf{k} \cdot \mathbf{r} - \omega t)\}] \tag{8.191^2}$$

The average magnitude of **S** over a period is, thus, given by

$$I(\omega) \equiv \left| \frac{1}{(2\pi/\omega)} \int_0^{2\pi/\omega} \mathbf{S}\,dt \right| = (\omega^2/2\pi c) \mid \mathbf{A}_0 \mid^2. \tag{8.192}$$

This is the intensity associated with the plane wave (8.187). Substituting from Eqs. (8.190) and (8.192) in (8.181a), we get,

$$w_{i \to f}^{(a)} = \frac{8\pi c}{\hbar^2} \mid \mathcal{M}_{if} \mid^2 \frac{I(\omega)}{\omega^2} \cdot \frac{\sin^2\left\{\frac{1}{2}(\omega - \omega_{fi})t\right\}}{(\omega - \omega_{fi})^2}, \tag{8.193}$$

where, $$\mathcal{M} = -(e/mc)\sum_j (\mathbf{a} \cdot \hat{\mathbf{p}}_j) \exp(-i\mathbf{k} \cdot \mathbf{r}_j);$$

$$\mathbf{A}_0 = \mid \mathbf{A}_0 \mid \mathbf{a}. \tag{8.194}$$

When the transition is between discrete levels, ω_{fi} would be a fixed quantity. Then, $w_{i \to f}^{(a)}$ considered as a function of ω will be a curve similar to the one in Fig. 8.11(b), with a sharp peak at $\omega = \omega_{fi}$. The total transition probability is obtained by integrating expression (8.193) over the range $\Delta\omega$ of frequencies:

$$W_{i \to f}^{(a)} = \int_{\Delta\omega} w_{i \to f}^{(a)} d\omega.$$

Just as in the case of Eq. (8.170^1), the limits of integration could be extended to $-\infty$ and $+\infty$ without appreciably affecting the value of the integral. Also, the factor

$I(\omega)/\omega^2$, being a slowly varying function of ω compared with the other factor involving ω, can be replaced by its value at $\omega = \omega_{fi}$. Thus,

$$W_{i-f}^{(a)} = \frac{8\pi c}{\hbar^2} |M_{if}|^2 \frac{I(\omega_{fi})}{\omega_{fi}^2} \cdot \int_{-\infty}^{+\infty} \frac{\sin^2(\omega'/2)t}{(\omega')^2} d\omega',$$

$$= \frac{4\pi^2 c}{\hbar^2 \omega_{fi}^2} |\mathcal{M}_{if}|^2 I(\omega_{fi}) t. \tag{8.194}$$

Therefore, the transition probability per unit time for a transition upward, in energy (absorption) between two discrete levels of the atom is given by,

$$\frac{dW_{i\to f}^{(a)}}{dt} \equiv (w_d^{(a)})_{i\to f} = \left(\frac{4\pi^2 c}{\hbar^2 \omega_{fi}^2}\right) |\mathcal{M}_{if}|^2 I(\omega_{fi}). \tag{8.195^1}$$

Similarly, the transition rate for a downward transition (induced emission) is,

$$(w_d^{(e)})_{i\to f} = \left(\frac{4\pi^2 c}{\hbar^2 \omega_{if}^2}\right) |\mathcal{M}_{fi}|^2 I(\omega_{if}). \tag{8.196^1}$$

Eqs. (8.195^1) and (8.196^1) give the rate of transitions induced by a radiation field consisting of an incoherent[29] mixture of harmonic waves of different frequencies, propagated in the direction **k** and *polarized*[30] in the direction **a**.

These equations show that

$$(w_d^{(a)})_{i\to f} = (w_d^{(e)})_{f\to i}. \tag{8.197}$$

That is, the probability that the field induces the transition $|u_i\rangle \to |u_f\rangle$ is the same as the probability that the field induces the inverse transition $|u_f\rangle \to |u_i\rangle$. This is known as the *principle of detailed balance*. In spite of this, the intensity of the stimulated emission between two atomic levels is, normally, much less than that of the reverse process. This is because of the greater initial number of atoms in the lower state (under normal circumstances). The stimulated emission can be, however, made to predominate over the absorption process by achieving an inversion of the normal population of the two levels, as in the case of masers and lasers.

Dipole Transitions

In Eqs. (8.195^1) and (8.196^1) the matrix elements of $\hat{\mathcal{M}}$ are the only quantities that depend on the nature of the atom. For the atomic case, the wavelength of the radiation, $\lambda \sim 10^{-5}$ cm, while r is of the order of atomic dimensions (10^{-8} cm). Thus, $\mathbf{k}\cdot\mathbf{r} \sim 10^{-3}$ ($k = 2\pi/\lambda$) and, therefore, the series,

$$\exp(-i\mathbf{k}\cdot\mathbf{r}) = 1 - i\mathbf{k}\cdot\mathbf{r} + \frac{1}{2}(-i\mathbf{k}\cdot\mathbf{r})^2 + \ldots, \tag{8.198}$$

---

29. The incoherence means that there is no particular phase-correlation between the different harmonic components.
30. By convention, the direction of the polarization of an electromagnetic field is defined as the direction of the electric vector.

converage very rapidly. As a result, it is a good approximation to replace the factor exp $(-i\mathbf{k}\cdot\mathbf{r}_j)$ in Eq. (8.194) by unity. Then, we have,

$$\hat{\mathcal{M}} \approx -\frac{e}{mc}\sum_{j=1}^{z}(\mathbf{a}\cdot\hat{\mathbf{p}}_j), \tag{8.199}$$

and

$$\mathcal{M}_{if} = -\frac{e}{mc}\sum_{j=1}^{z}\mathbf{a}\cdot(\hat{\mathbf{p}}_j)_{if}, \tag{8.200^1}$$

where $\mathbf{p}_j$ is the matrix corresponding to $\hat{\mathbf{p}}_j$.

Now,

$$(\mathbf{p})_{if} = m\left(\frac{d\mathbf{r}}{dt}\right)_{if} = (m/i\hbar)\,[\hat{\mathbf{r}}, \hat{H}]_{if}$$

$$= \frac{m}{i\hbar}\{\langle u_i \mid \hat{\mathbf{r}}\hat{H} \mid u_f\rangle - \langle u_i \mid \hat{H}\hat{\mathbf{r}} \mid u_f\rangle\}$$

$$= (m/i)\,[(E_f - E_i)/\hbar]\,\langle u_i \mid \hat{\mathbf{r}} \mid u_j\rangle$$

$$= -im\,\omega_{fi}(\mathbf{r})_{if}. \tag{8.201}$$

Substituting in (8.200^1) from (8.201), we get,

$$\mathcal{M}_{if} = i(\omega_{if}/c)\,(\mathbf{a}\cdot\mathbf{D}_{if}), \tag{8.200^2}$$

where,

$$\mathbf{D} = \sum_{j=1}^{z} e\hat{\mathbf{r}}_j, \tag{8.202}$$

is the *electric dipole moment operator*. The approximation (8.199) is, for this reason, known as the *dipole approximation*.

The transition probability per unit time for *electric dipole transitions* in the atom is, thus, obtained by substituting (8.200^2) in Eqs. (8.195^1) and (8.196^1):

$$(w_d^{(a)})_{i\to f} = \frac{4\pi^2}{\hbar^2 C}\cdot I(\omega_{fi})\,|\,(\mathbf{a}\cdot\mathbf{D}_{if})\,|^2, \tag{8.195^2}$$

$$(w_d^{(e)})_{i\to f} = \frac{4\pi^2}{\hbar^2 C}\cdot I(\omega_{if})\,|\,(\mathbf{a}\cdot\mathbf{D}_{fi})\,|^2. \tag{8.196^2}$$

When the electromagnetic field is unpolarized, then $|\,\mathbf{a}\cdot\mathbf{D}_{if}\,|^2$ in (8.195^2) should be replaced by its average value over the different possible relative orientations of $\mathbf{a}$ and $\mathbf{D}_{if}$. If θ represents the angle between $\mathbf{a}$ and $\mathbf{D}_{if}$, the average of $|\,\mathbf{a}\cdot\mathbf{D}_{if}\,|^2$ is given by,

$$|\,\mathbf{a}\cdot\mathbf{D}_{if}|^2_{\text{average}} = \frac{1}{4\pi}\int |\,\mathbf{a}\cdot\mathbf{D}_{if}\,|^2\,d\Omega$$

$$= \frac{1}{4\pi} |\mathbf{D}_{if}|^2 \int_{-1}^{+1} \cos^2\theta d(\cos\theta) \int_0^{2\pi} d\phi$$

$$= \frac{1}{3} |\mathbf{D}_{if}|^2 . \tag{8.203}$$

Here, $d\Omega \equiv \sin\theta d\theta d\phi$, is an element of solid angle. Thus, for transitions induced by an unpolarised field, we have,

$$(w_d^{(a)})_{i\to f} = \frac{4\pi^2}{3\hbar^2 c} I(\omega_{fi}) | < u_i | \hat{\mathbf{D}} | u_f > |^2, \tag{8.195^3}$$

$$(w_d^{(e)})_{i\to f} = \frac{4\pi^2}{3\hbar^2 c} I(\omega_{if}) | < u_f | \hat{\mathbf{D}} | u_i > |^2 . \tag{8.196^3}$$

Selection Rules

From the properties of the dipole operator $\hat{\mathbf{D}}$ (Eq. (8.202)), we could deduce certain rules that are to be satisfied if the matrix element $< u_f | \hat{\mathbf{D}} | u_i >$, and hence the transition probability $(w_d)_{i\to f}$ are not to be zero. These rules are summarised below:

(i) The state $| u_f >$ should differ from $| u_i >$ in the state of only one electron. That is, transitions in which more than one electron changes state are forbidden (Such transitions can occur only in the higher order approximations of the perturbation theory).

(ii) If J_i, M_i and J_f, M_f are the angular momentum quantum numbers associated with $| u_i >$ and $| u_j >$ respectively, then,

$$\Delta J \equiv J_f - J_i = 0, \pm 1;$$

but no $J_i = 0$ to $J_f = 0$ transition.

(iii) The states $| u_i >$ and $| u_f >$ should have opposite parties.

Problem 8.15: Deduce the above selection rules.

In the foregoing treatment of the atomic radiation, we have treated the atom quantum mechanically and the electromagnetic field classically. For this reason, the theory falls under the category of *semi-classical* theories. However, a satisfactory account of radiation can be obtained only within the framework of a theory in which the electromagnetic field also is subjected to the rules of quantization. Such a *quantum theory of radiation* yields a result[31] which is in agreement with formula (8.195^1) in the case of absorption, but which differs from the formula (8.196^1) for emission by an additional term which does not vanish in the absence of the field. The additional term is interpreted as representing the probability for *spontaneous emission* (see footnote 27) by the atom.

31. See, for example, Schiff, L.I., op. cit. Eqs. (57.26) and (57.28).

Problem 8.16: In Problem 8.14, let ω be less than the ionizing frequency of the atom. In this case, calculate the probability per unit time for the atom to make a transition to an upper level. What is the polarity of the transition?

8.4C. Coulomb Excitation

As an example of a time-dependent perturbation that is neither constant nor harmonic in time, we consider the process of coulomb excitation of nuclei in which a beam of charged particles, such as protons, collides with a nucleus, exciting the latter from an initial state $|u_i>$ to a final state $|u_f>$. Here also, we will adopt a semi-classical approach, treating states of the target (the nucleus) quantum mechanically and the orbit of the *projectile* (the charged particle) classically[32].

Let Ze = charge of the nucleus,
R = radius of the nucleus,
J_i, M_i = the angular momentum quantum numbers characterising the state $|u_i>$,
J_f, M_f = the angular momentum quantum numbers characterising the state $|u_f>$,
E_i, E_f = the energies of the state $|u_i>$ and $|u_f>$.
$E = \frac{1}{2}Mv^2$, the energy of the projectile (assumed to be a proton) in the C.M. system.
Ω_i, Ω_f = the initial (incident) and the final (scattered) directions of the proton in the C.M. system.
θ = the scattering angle (the angle between the directions Ω_i and Ω_f).

The Hamiltonian of the system can be written as

$$\hat{H} = \hat{H}_N + \hat{H}_p + \hat{H}_pN, \tag{8.204^1}$$

where, $\hat{H}_N$ and $\hat{H}_p$, respectively, refer to the nucleus and the proton while $\hat{H}_{pN}$ represents the interaction between the two:

$$\hat{H}_N |u_i> = E_i |u_i>;$$

$$\hat{H}_N |u_f> = E_f |u_f>. \tag{8.205}$$

When the energy of the proton is small enough, the proton will not penetrate the nucleus. In this case the proton-nucleus interaction will be purely electrostatic (Coulomb) and is given by

$$\hat{H}_{pN} = e^2 \sum_{j=1}^{Z} \frac{1}{|\mathbf{r}_p(t) - \mathbf{r}_j|}, \tag{8.206^1}$$

where, $\mathbf{r}_p(t)$ is the position of the proton w.r.t. the centre of the nucleus and $\mathbf{r}_j$ that of the jth proton in the nucleus. For large values of r_p (that is, $r_p \gg R$), we have,

---

32. The treatment here largely follows the one given in Messiah, A., *Quantum Mechanics*, Chapter XVII, Section 3.

$$\hat{H}_{pN} \approx \hat{U}(r_p) = e^2 \sum_{j=1}^{Z} \frac{1}{r_p(t)} = \frac{Ze^2}{r_p(t)}. \qquad (8.206^2)$$

The Hamiltonian of the system is, then, $\hat{H} = \hat{H}_N + \hat{H}'_p$, where $\hat{H}_N$ involves the co-ordinates of only the nucleus while $\hat{H}'_p$ contains the co-ordinates of only the projectile. The motion, thus, separates out into the individual motions of the nucleus and the proton, the latter in the coulomb field provided by the former. The problem reduces to that of elastic scattering (no transfer of energy and, hence, no internal excitation of the nucleus) by a potential (Section 7.1). The differential scattering cross-section is given by the Rutherford formula (Eq. (7.116)):

$$\frac{d\sigma}{d\Omega} = \left(\frac{Ze^2}{2Mv^2}\right)^2 \frac{1}{\sin^4(\theta/2)} = \frac{a^2}{4\sin^4(\theta/2)}, \qquad (8.207)$$

where $a = (Ze^2/2E)$. The length $r_0 = 2a$, is the distance of closest approach[33] of the proton to the centre of the nucleus, and the condition $(E = \hbar^2 k^2/2M)$,

$$kr_0 \gg 1, \qquad (8.208)$$

is required for the validity of Eq. (8.207) [See the discussion on the criterion for the validity of the Born Approximation, Section 7.4].

As the proton approaches the nucleus, $\hat{H}_{pN}$ will depart more and more from $\hat{U}(r_p)$. As long as $r_p > R$, the difference between $\hat{H}_{pN}$ and $\hat{U}(r_p)$ can be represented by,

$$\hat{V}(t) = e^2 \sum_{j=1}^{Z} \left\{ \frac{1}{|\mathbf{r}_p(t) - \mathbf{r}_j|} - \frac{1}{r_p} \right\}, \qquad (8.209^1)$$

and $\hat{H}$ can be written as

$$\hat{H} = \hat{H}_0 + \hat{V}(t), \qquad (8.209^2)$$

with

$$|\hat{V}(t)| \ll |\hat{H}_0|, \qquad (8.210)$$

where,

$$\hat{H}_0 = \hat{H}_N + \hat{H}_p + \hat{U}(r_p). \qquad (8.211)$$

Thus, $\hat{V}(t)$ can be treated as a time-dependent perturbation (note that $\hat{V}(t)$ is neither constant nor harmonic in time) that can cause transitions between the eigenstates of $\hat{H}_0$ in accordance with Eq. (8.165^2).

Now, the eigenvectors of $\hat{H}_0$ would be products of the eigenvectors of $\hat{H}_N$ and the eigenvectors of $\hat{H}'_p = \hat{H}_p + \hat{U}(r_p)$. It is here that we resort to the semi-classical approach. Instead of using the eigenstates of $\hat{H}'_p$, and treating $\hat{V}(t)$ as a perturbation that causes transitions between such eigenstates (along with transitions between the eigenstates of $\hat{H}_N$), we use the classical orbit of the proton in the Coulomb field represented by $\hat{U}(r_p)$ and regard $\hat{V}(t)$ as a perturbation that

33. The point of closest approach is a classical turning point defined by $E = U_{\text{eff}}(r) = \frac{Ze^2}{r} + \frac{l(l+1)\hbar^2}{2Mr^2}$ [See Eq. (7.47)]. The minimum value of r for which this condition is satisfied is obtained when the angular momentum l (and, hence, the impact parameter) is zero.

causes transitions only between the eigenstates of the nucleus. Such an approach neglects the energy loss that the proton necessarily suffers in the process of exciting the nucleus from a lower to a higher energy state[34], and would, therefore, be justified only if,

$$\Delta E = (E_f - E_i) \ll E. \tag{8.212}$$

This is in addition to the condition (8.208) which is required for treating the proton-orbit classically.[35]

According to Eq. (8.165^2), the probability per unit time for the nucleus to make a transition from the state $|u_i> \equiv |J_i M_i>$ to the state $|u_f> \equiv |J_f M_f>$ under the influence of $\hat{V}(t)$, is given by

$$w_{i\to f} = \frac{1}{\hbar^2}\left| \int_{-\infty}^{+\infty} <J_f M_f|\hat{V}(t)|J_i M_i> \exp\left[(i/\hbar)\Delta E t\right] dt \right|^2, \tag{8.213^1}$$

where, the initial and the final times are taken, respectively, as long before scattering and long after scattering. The differential scattering cross-section for the inelastic process is the product of $w_{i\to f}$ and the cross-section (8.207) for the elastic process:

$$\frac{d\sigma_{i\to f}}{d\Omega} = \left(\frac{a^2}{4\sin^2(\theta/2)}\right) w_{i\to f}. \tag{8.214^1}$$

This gives the probability per unit time that the proton is scattered from the initial direction Ω_i to the final direction Ω_f and that, at the same time, the nucleus has made a transition from $|u_i>$ to $|u_f>$.

Now,
$$\frac{1}{|\mathbf{r}_p - \mathbf{r}_j|} = \frac{1}{r_p}(1 - 2x\cos\Theta_j + x^2)^{-1/2},$$

where $x = r_j/r_p$ and Θ_j is the angle between $\mathbf{r}_p$ and $\mathbf{r}_j$. Then, from Eq. (E.28) and the spherical harmonic addition theorem (5.152), we get,

$$\frac{1}{|\mathbf{r}_p - \mathbf{r}_j|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} (-1)^m r_j^l\, Y_{lm}(\theta_j, \phi_j) r_p^{-l-1}\, Y_{l-m}(\theta_p, \phi_p). \tag{8.215a}$$

Also,
$$\frac{1}{r_p} = \frac{1}{r_p}\left(\frac{r_j}{r_p}\right)^0 P_0(\cos\theta_j). \tag{8.215b}$$

Thus,
$$\hat{V}(t) = e^2 \sum_{j=1}^{Z}\left[\frac{1}{|\mathbf{r}_p(t) - \mathbf{r}_j|} - \frac{1}{r_p}\right]$$

$$= \sum_{j=1}^{\infty} \sum_{m=-l}^{+l} (-1)^m \hat{Q}_m^{(l)}\, T_{-m}^{(l)}, \tag{8.209^2}$$

---

34. This process of nucleus excitation is termed *Coulomb excitation* in order to distinguish it from other processes in which the short-range, and more powerful, nuclear force comes into play.
35. Formula (8.207), which was derived in chapter 7 in the Born approximation, is also the classical Rutherford formula [See, for example, Goldstein, H., *Classical Mechanics* (Addison-Wesley, Massachussetts, 1959), Eq. (3.68)].

where,
$$\hat{Q}_m^{(l)} = e \sum_{j=1}^{Z} \hat{r}_j^l Y_{lm}(\theta_j, \phi_j), \tag{8.216a}$$

and[36]

$$T_m^{(l)} = \left(\frac{4\pi}{2l+1}\right) e r_p^{-l-1} Y_{lm}(\theta_p, \phi_p). \tag{8.216b}$$

The $\hat{Q}_m^{(l)}$ are the components of the *electric multipole moment operator* $\hat{Q}^{(l)}$ of order l for the nucleus. $\hat{Q}^{(l)}$ represents the part of $\hat{V}(t)$ that acts on the co-ordinates of the nucleus. That is,

$$< J_f M_f | \hat{V}(t) | J_i M_i > = \sum_{l,m} < J_f M_f | \hat{Q}_m^{(l)} | J_i M_i > (-1)^m T_{-m}^{(l)},$$

so that, Eq. (8.213[1]) becomes,

$$w_{i \to f} = | \sum_{l=1}^{\infty} \sum_{m=-l}^{+l} (-1)^m S_{-m}^{(l)} < J_f M_f | \hat{Q}_m^{(l)} | J_i M_i > |^2 \tag{8.213[2]}$$

where,
$$S_{-m}^{(l)} = \frac{1}{\hbar} \int_{-\infty}^{+\infty} T_{-m}^{(l)}(t) \exp\{i(\Delta E/\hbar)t\} dt. \tag{8.217}$$

$S_{-m}^{(l)}$ can be evaluated from a knowledge of the classical path of the proton. Since $\hat{Q}^{(l)}$ is a spherical tensor of rank l (see Section 5.6, Eq. (5.170[1])) and of parity $(=1)^l$, for given states $| J_i M_i >$ and $| J_f M_f >$, we have,

$$| J_f - J_i | \leq l \leq J_f + J_i; m = M_f - M_i, \tag{8.218a}$$

$$\pi_i \pi_j = (-1)^l. \tag{8.218b}$$

These are the *selection rules* that have to be satisfied for a nonvanishing $w_{i \to f}$. The parity selection rule (8.218b) restricts the summation in (8.213[2]) either to even values of l or to odd values of l. Moreover, we see from Eqs. (8.216a, b) and (8.213[2]), that a transition of order $(l+2)$ is less probable by a factor of the order of $(R/r_0)^2$ than a transition of order l. Since (R/r_0) is small, we need retain only the lowest value of l, say $l_0 = | J_f - J_i |$ or $| J_f - J_i | + 1$, consistent with (8.218b). Thus,

$$w_{i \to f} \approx | < J_f M_f | \hat{Q}_m^{(l_0)} | J_i M_i > |^2 \cdot | S_{-m}^{(l_0)} |^2, (m = M_f - M_i)$$

$$= \frac{1}{2J_f + 1} \left(C_{M_i m M_f}^{J_i l_0 J_f} \right)^2 | < J_f \| Q^{(l_0)} \| J_i > |^2 \cdot | S_{-m}^{(l_0)} |^2, \tag{8.213[3]}$$

where, the Wigner-Eckart theorem (Eq. (5.170[2])) has been used.

In (8.123[3]), the nucleus is polarized both in its initial state and final state (since it has a definite angular momentum projection along the z-axis). But when the target nucleus is unpolarized initially, and the polarization of the final state is not observed, we should average over the $(2J_i + 1)$ orientations of the initial state and sum over the $(2J_f + 1)$ orientations of the final state. Thus,

$$w_{i \to f}(\text{unpolarized}) = \frac{1}{(2J_i + 1)} \sum_{M_i = -J_i}^{+J_i} \sum_{M_f = -J_f}^{+J_f} w_{i \to f}(\text{polarized})$$

---

36. Unlike $\hat{Q}_m^{(l)}$, $T_m^{(l)}$ is regarded as a classical quantity (that is, not as an operator).

$$= \frac{|<J_f||\hat{Q}^{(l_0)}||J_i>|^2}{(2J_i+1)(2J_f+1)} \cdot \sum_{M_i M_f} |S_{-m}^{(l_0)}|^2 \left(C_{M_i m M_f}^{J_i l_0 J_f}\right)^2$$

$$= \frac{|<J_f||\hat{Q}^{(l_0)}||J_i>|^2}{(2J_i+1)(2l_0+1)} \Sigma_m |S_{-m}^{(l_0)}|^2, \qquad (8.213^4)$$

since $$\sum_{M_i M_f} |S_{-m}^{(l_0)}|^2 \left(C_{M_i m M_f}^{J_i l_0 J_f}\right)^2 = \sum_m |S_{-m}^{(l_0)}|^2 \sum_{M_i} \left(C_{M_i m M_f}^{J_i l_0 J_f}\right)^2$$

$$= \sum_m |S_{-m}^{(l_0)}|^2 \frac{2J_f+1}{2l_0+1}, \text{ (by Eq. (5.84d)).}$$

Substituting from (8.213^4) into (8.214^1), we get,

$$\frac{d\sigma_{i\to f}}{d\Omega} = \frac{a^2 |<J_f||\hat{Q}^{(l_0)}||J_i>|^2}{4(2J_i+1)(2l_0+1)} \cdot \left\{\sin^{-4}(\theta/2) \cdot \Sigma_m |S_m^{(l_0)}|^2\right\}. \qquad (8.214^2)$$

[Note that $$\sum_m |S_{-m}^{(l_0)}|^2] = \sum_{m'} |S_{m'}^{(l_0)}|^2].$$

This formula gives the differential cross-section for the Coulomb excitation of nuclei. In the case of nuclei, the lowest multipole of interest is of order $2(l_0 = 2)$. Also, the matrix element $<J_f||\hat{Q}^{(2)}||J_i>$ is appreciable only when the nuclear states arise from the collective motion of the nucleus, rather than from the promotion of one or two nucleons to a higher shell model orbit. The phenomenon of coulomb excitation is, thus, a valuable tool in determining the spin, parity and mean life of collective states in nuclei.

8.5 SUDDEN AND ADIABATIC APPROXIMATIONS

These approximations deal with the modification of the state of a system when the Hamiltonian of the system is changed from $\hat{H}_0$ at $t = t_0$ to $\hat{H}_1$ at $t = t_1$. This modification will depend critically on the time $T = (t_1 - t_0)$ during which the change of the Hamiltonian takes place. The limiting cases of very small T (sudden change) and of very large T (adiabatic change) are comparatively simple. The approximation methods dealing with these cases respectively are known as the *sudden approximation* and the *adiabatic approximation.* A description of these approximations, along with a discussion of the criteria for their validity, is the subject matter of this section.

8.5A. Sudden Approximation

It is convenient to define,

$$s = \frac{t - t_0}{T}, \qquad (8.219)$$

and parametrise $\hat{H}$ in terms of s. Then,

$$\hat{H}_0 \equiv \hat{H}(s=0) = \hat{H}(0); \hat{H}_1 = \hat{H}(1).$$

Also, let $\hat{U}(t, t_0) = \hat{U}_T(s)$.

Eq. (4.13), then, takes the form,

$$\hat{U}_T(s) = \hat{1} - (i/\hbar)T \int_0^s \hat{H}(s')\hat{U}_T(s')ds',$$

or,

$$U_T(1) = \hat{1} - (i/\hbar)T \int_0^1 \hat{H}(s)\hat{U}_T(s)ds. \qquad (8.220^1)$$

Thus,

$$\text{Lim}_{T\to 0}\hat{U}_T(1) = \hat{1}, \qquad (8.221)$$

so that

$$\psi(t_1) = \hat{U}(t_1, t_0)\psi(t_0) = \hat{U}_T(1)\psi(t_0) = \psi(t_0) \qquad (8.222)$$

That is, the system does not change state eventhough the Hamiltonian has changed. The sudden approximation consists in setting $\hat{U}_T(1) = \hat{1}$ when T is small but not zero. A measure of the error involved in this approximation is given by the probability w of finding the system in states other than the initial state $|u_i> \equiv \psi(t_0)$.

Let $|\psi_f>$ represent the state of the system at $t = t_1$. In general, $|\psi_f>$ would be a linear superposition of the eigenstates $\{|u_k>\}$ of $\hat{H}_0$. The part of $|\psi_f>$ that is orthogonal to the initial state $|u_i>$ is given by

$$|\psi_i> = |\psi_f> - |u_i><u_i|\psi_f> = \hat{Q}_i|\psi_f>, \qquad (8.223)$$

where,

$$\hat{Q}_i = \hat{1} - |u_i><u_i| = \hat{1} - \hat{\pi}_i, \qquad (8.224)$$

is the *projection operator* that projects onto the subspace orthogonal to $|u_i>$. Then,

$$w = <\psi_f|\hat{Q}_i|\psi_f>, \qquad (8.225^1)$$

But

$$|\psi_f> \equiv \psi(t_1) = \hat{U}_T(1)|u_i>, \qquad (8.222^2)$$

so that,

$$w = <u_i|\hat{U}_T^{\dagger}(1)\hat{Q}_i\hat{U}_T(1)|u_i>. \qquad (8.225^2)$$

Now, by an iterative procedure identical to the one leading to Eqs. (8.159^3) and (8.160a), we convert (8.220^1) to the form,

$$\hat{U}_T(1) = \sum_{n=0}^{\infty} \hat{U}_T^{(n)}(1), \qquad (8.220^2)$$

with

$$\hat{U}_T^{(n)}(1) = (-i/\hbar)^n T^n \int_0^1 ds_1 \int_0^{s_1} ds_2 \ldots \int_0^{s_{n-1}} ds_n \cdot \hat{H}(s_1)\hat{H}(s_2)\ldots\hat{H}(s_n),$$

(8.226)

where,

$$1 > s_1 > s_2 > \ldots > s_{n-1}.$$

Substituting in (8.225^2) from (8.220^2) and (8.226), we get,

$$w = < u_i \mid \hat{Q}_i \mid u_i > + (i/\hbar)T\{< u_i \mid (\overline{\hat{H}}\,\hat{Q}_i - \hat{Q}_i\,\overline{\hat{H}})u_i >\}$$

$$+ (T^2/\hbar^2)[< u_i \mid \overline{\hat{H}}\,\hat{Q}_i\,\overline{\hat{H}} \mid u_i >$$

$$- \{< u_i \mid \hat{Q}_i \int_0^1 ds_1 \int_0^{s_1} ds_2 \hat{H}(s_1)\hat{H}(s_2) \mid u_i >$$

$$+ < u_i \mid \int_0^1 ds_1 \int_0^{s_1} ds_2 \hat{H}(s_1)\hat{H}(s_2)\hat{Q}_i \mid u_i >\}]$$

$$+ \ldots,$$

where,

$$\overline{\hat{H}} = \int_0^1 \hat{H}(s)ds = \frac{1}{T}\int_{t_0}^{t_1} \hat{H}(t)dt,$$

is the average of $\hat{H}$ over the interval T.

We have,

$$\hat{Q}_i \mid u_i > = 0 = < u_i \mid \hat{Q}_i.$$

Therefore,

$$w \approx (T^2/\hbar^2) < u_i \mid \overline{\hat{H}}\,\hat{Q}_i\,\overline{\hat{H}} \mid u_i > = (\Delta\overline{E}/\hbar)^2 T^2, \qquad (8.225^3)$$

where [c.f. Eq. (3.25a)],

$$(\Delta\overline{E})^2 = < u_i \mid \overline{\hat{H}}^2 \mid u_i > - < u_i \mid \overline{\hat{H}} \mid u_i >^2. \qquad (8.227)$$

$\Delta\overline{E}$ could be regarded as a measure of the average change in the energy of the system during the time T. Since T is small, we have, retained only the lowest power of T with a nonvanishing coefficient[37].

Thus, the condition for the validity of the sudden approximation, $w \ll 1$, becomes,

$$T \ll \hbar/\Delta\overline{E}. \qquad (8.228)$$

This result may be compared with relationship (8.169b) obtained in the case of the constant perturbation. The two results could be regarded as the two sides of the same coin if we interpret $(\hbar/\Delta\overline{E})$ as the *characteristic time* τ (the minimum time) associated with an energy change $\Delta\overline{E}$ that is accompanied by a change of state (Recall that change of energy can take place without change of state, as also change of state without energy change). Then, inequality (8.228) states that, if the energy of a system changes by the amount $\Delta\overline{E}$ in a time T which is very much less than the characteristic time associated with such a change of energy, then the

37. Note that $\mid u_i >$ should not be an eigenvector of $\overline{\hat{H}}$ for $\Delta\overline{E}$ not to vanish.

system does not change state. The other side would be that $\hbar/T$ is the characteristic (*maximum*) energy-change associated with a change of state that takes place during a time T, so that if $\Delta\bar{E} \gg \hbar/T$, there would be no change of state (no transition).

Sudden Reversal of a Magnetic Field

As an example of the application of the sudden approximation, let us consider what happens to an atom in a magnetic field when the direction of the field is suddenly reversed[38]. The following assumptions will be made:

(i) Atom is one for which the *LS*-coupling[39] is valid, so that a state of the atom, in the angular momentum representation, can be denoted by $|LSM_LM_S>$, where L and S represent the orbital angular momentum and spin while M_L and M_S represent their components along the z-axis.

(ii) The magnetic field **B** is strong enough to decouple **L** and **S**.

(iii) **B** is along the z-axis and varies linearly with time according to the formula,

$$\mathbf{B}(s) = \mathbf{B}_0(2s-1), \left(s = \frac{t-t_0}{T}\right). \tag{8.229}$$

The Hamiltonian of the atom is, then, given by (see Problem 8.10),

$$\hat{H}(s) = \hat{H}^{(0)} + \alpha(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}) - \frac{e}{2\mu c}(\hat{L}_z + 2\hat{S}_z)B(s). \tag{8.230}$$

That is,

$$\hat{H}_0 = \hat{H}^{(0)} + \alpha(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}) + \frac{eB_0}{2\mu c}(\hat{L}_z + 2\hat{S}_z), \tag{8.230a}$$

$$\hat{H}_1 = \hat{H}^{(0)} + \alpha(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}) - \frac{eB_0}{2\mu c}(\hat{L}_z + 2\hat{S}_z). \tag{8.230b}$$

Here, $\hat{H}^{(0)}$ is the Hamiltonian in the *LS*-coupling scheme and $\alpha(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}})$ represents the spin-orbit coupling. Since the average value of **B** is zero, we have,

$$\overline{\hat{H}} = \hat{H}^{(0)} + \alpha(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}), \tag{8.231}$$

and

$$(\overline{\hat{H}})^2 = (\hat{H}^{(0)})^2 + \alpha^2(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}})^2 + \hat{H}^{(0)}\alpha(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}) + \alpha(\hat{\mathbf{L}}\cdot\hat{\mathbf{S}})\hat{H}^{(0)},$$

---

38. This example, as well as the one in the case of the adiabatic approximation in Section 8.5B, is taken from A. Messiah, op. cit. Chapter XVII.

39. In *LS*-coupling, the orbital angular momenta $\mathbf{l}_i$ and spins $\mathbf{s}_i$ of the individual electrons are added separately to obtain the total orbital angular momentum $\mathbf{L} = \sum_{i=1}^{Z} \mathbf{l}_i$ and the total spin $\mathbf{S} = \sum_{i=1}^{Z} \mathbf{s}_i$. In the *jj-coupling*, on the other hand, $\mathbf{l}_i$ and $\mathbf{s}_i$ are added to obtain a total angular momentum $\mathbf{j}_i = \mathbf{l}_i + \mathbf{s}_i$ for each electron and, then, the $\mathbf{j}_i$ are added to yield a total angular momentum $\mathbf{J} = \sum_{i=1}^{Z} \mathbf{j}_i$, for the atom.

so that

$$(\Delta\overline{E})^2 = < LSM_L M_S \,|\, \overline{H}^2 \,|\, LSM_L M_S > - < LSM_L M_S \,|\, \overline{H} \,|\, LSM_L M_S >^2$$

$$= \alpha^2 \{ < (\hat{\mathbf{L}} \cdot \hat{\mathbf{S}})^2 > - < (\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}) >^2 \}. \qquad (8.232^1)$$

Now,

$$(\hat{\mathbf{L}} \cdot \hat{\mathbf{S}}) = \frac{1}{2}(\hat{L}_+ \hat{S}_- + \hat{L}_- \hat{S}_+) + \hat{L}_z \hat{S}_z,$$

where $\hat{L}_\pm$ and $\hat{S}_\pm$ are the angular momentum raising and lowering operators [Eq. (5.14)]. It follows that

$$(\Delta\overline{E})^2 = \alpha^2 \left[< \{ \frac{1}{2}(\hat{L}_+ \hat{S}_- + \hat{L}_- \hat{S}_+) \}^2 > - < \frac{1}{2}(\hat{L}_+ \hat{S}_- + \hat{L}_- \hat{S}_+) >^2 \right].$$

$$= \frac{1}{4}\alpha^2 \hbar^4 [2\{L(L+1) - M_L^2\}\{S(S+1) - M_s^2\} - 2M_L M_S], \qquad (8.232^2)$$

where, Eqs. (5.31a, b) have been used.

The maximum value of the quantity in the square bracket in Eq. (8.232^2) is $2L(L+1)S(S+1)$ (corresponding to $M_L = M_S = 0$). Hence $\Delta\overline{E}$ is of the order of $\alpha\hbar^2$, and the condition (8.228) for the validity of the sudden approximation becomes:

$$T \ll \frac{1}{\alpha\hbar}. \qquad (8.233)$$

Now, $\alpha\hbar^2$ is of the order of the spin-orbit splitting [see Eqs. (8.153a, b)], so that $1/\alpha\hbar$ is the characteristic time associated with the spin-orbit splitting. Thus, the condition for the atom not to change state during the reversal of the magnetic field is that this reversal should take place in a time which is small compared with the time characteristic of the spin-orbit splitting. This result could be understood as follows: It is the spin-orbit term in (8.230) that causes transition between states, as the other terms in the equation are diagonal in the representation defined by the basis vectors $|\, LSM_L M_S >$. But the spin-orbit term can cause transitions only when it dominates over the magnetic term in (8.230). Therefore, if the reversal of the magnetic field, where the magnetic energy dominates over the spin-orbit energy both before and after the reversal, is accomplished within a short period compared with the time required for the spin-orbit interaction to effect a transition, no transition will take place.

8.5B Adiabatic Approximation

The Adiabatic Theorem: This states that, if the system is initially in an eigenstate of $\hat{H}_0 \equiv \hat{H}(t_0)$, it would have passed, in the limit $T \to \infty$, into the eigenstate of $\hat{H}_1 \equiv \hat{H}(t_1)$, that derives from the eigenstate of $\hat{H}_0$ by continuity[40].

The adiabatic approximation consists in assuming the validity of the adiabatic theorem when T is large but not infinite.

---

40. For a proof of the theorem, see Messiah, A. op. cit.

The error involved in this approximation can be estimated by calculating the probability of finding the system at time t_1 in a state different from the one indicated by the adiabatic theorem. Now, the implication of the adiabatic theorem is that the change of state, entailed by the time-dependence of the Hamiltonian, is a continuous process (that is, does not involve discrete jumps). Therefore, it would be meaningful to talk of stationary eigenfunctions of the *instantaneous Hamiltonian*:

$$\hat{H}(t)u_n(t) = E_n(t)u_n(t), \tag{8.234}$$

where, the u_n are assumed to be orthonormal, discrete and non-degenerate.

Continuity, then, means that if the system is in the eigenstate $u_m(t_0)$ with energy $E_m(t_0)$ at time t_0, it should be in the eigenstate $u_m(t)$ with energy $E_m(t)$ at time $t > t_0$. If $\hat{H}$ were independent of time $u_m(t)$ would be the same as $u_m(t_0)$ (that is, the state would be truly stationary), whereas here $u_m(t)$ could be quite different from $u_m(t_0)$ eventhough the former is evolved from the latter (see the example given at the end of this Section). In a discontinuous change, on the other hand, the system would be found in an eigenstate $u_k(t)$ of $\hat{H}(t)$, that does not evolve from $u_m(t_0)$. Thus, the condition for the validity of the adiabatic approximation is that the probability for finding the system in the state $u_k(t)$ be small if the system was initially in the state $u_m(t_0)$. A derivation of this probability is given below:

The fact that the system is likely to be found in states other than that prescribed by the adiabatic approximation, can be expressed by saying that the state vector $\Psi(t)$ of the system at time $t > t_0$ is a superposition of the stationary eigenstates of $\hat{H}(t)$. That is,

$$\Psi(t) = \sum_n a_n(t)u_n(t) \exp\left[-(i/\hbar)\int_{t_0}^{t} E_n(t')dt'\right] \tag{8.235}$$

where (see Postulate III, Chapter 3) $|a_n(t)|^2$ is the probability that the system is in the eigenstate $u_n(t)$ at time t. (Note that, if $\hat{H}$ were independent of time , both a_n and E_n in (8.235) would have been independent of time). Now, $\Psi(t)$ is the solution of the time-dependent Schrödinger equation,

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}(t)\Psi(t). \tag{8.236}$$

Substituting for Ψ in (8.236) from (8.235) and making use of Eq. (8.234) we get,

$$\sum_n(\dot{a}_n u_n + a_n\dot{u}_n) \exp\left[-(i/\hbar)\int_{t_0}^{t} E_n(t')dt'\right] = 0, \tag{8.237}$$

where, the dot denotes differentiation with respect to time. We now take the scalar product of this equation with $u_k(t)$. In view of the orthonormality of the u_n's, we have,

$$\dot{a}_k = -\sum_n a_n(t)\langle u_k(t) \mid \dot{u}_n(t)\rangle \exp\left(i\int_{t_0}^{t}\omega_{kn}(t')dt'\right). \qquad (8.238^1)$$

The product,

$$\langle u_k \mid \dot{u}_n\rangle \equiv \langle u_k \left| \frac{\partial u_n}{\partial t}\right\rangle$$

occurring here may be evaluated as follows: We have,

$$\frac{\partial}{\partial t}\{\langle u_k \mid \hat{H} \mid u_n\rangle\} = \langle \frac{\partial u_k}{\partial t} \mid \hat{H} \mid u_n\rangle + \langle u_k \mid \hat{H} \mid \frac{\partial u_n}{\partial t}\rangle + \langle u_k \left| \frac{\partial \hat{H}}{\partial t}\right| u_n\rangle.$$

Using Eq. (8.234) and the relationship,

$$0 = \frac{\partial}{\partial t}\langle u_k \mid u_n\rangle = \langle \dot{u}_k \mid u_n\rangle + \langle u_k \mid \dot{u}_n\rangle, \qquad (8.239)$$

we get,

$$\langle u_k \mid u_n\rangle \frac{\partial E_n}{\partial t} = (E_k - E_n)\langle u_k \mid \dot{u}_n\rangle + \langle u_k \left| \frac{\partial \hat{H}}{\partial t}\right| u_n\rangle.$$

Thus,

$$\langle u_k \mid \dot{u}_n\rangle = \frac{-1}{\hbar\omega_{kn}}\langle u_k \left| \frac{\partial \hat{H}}{\partial r}\right| u_n\rangle, k \neq n. \qquad (8.240)$$

From (8.239), we have,

$$< u_n \mid \dot{u}_n > = - < \dot{u}_n \mid u_n > = - < u_n \mid \dot{u}_n >^* .$$

Therefore, $< u_n \mid \dot{u}_n >$ is either zero or pure imaginary. It is possible to make it zero by a suitable choice of the phase[41] of u_n. Hence (8.238^1) becomes,

$$\frac{da_k(t)}{dt} = \sum_{n \neq k} \frac{a_n(t)}{\hbar\omega_{kn}(t)}\langle u_k(t) \left| \frac{\partial \hat{H}}{\partial t}\right| u_n(t)\rangle \cdot \exp\left(i\int_{t_0}^{t}\omega_{kn}(t')dt'\right). \qquad (8.238^2)$$

In order to integrate this expression, we replace $a_n(t)$, $\omega_{kn}(t)$ and $< u_k \mid \partial\hat{H}/\partial t \mid u_n >$ on the right hand side, by their values at $t = t_0$. Such a procedure would be justified in view of the slow variation of these quantities with time. Further, we assume that the system is in the state u_m at $t = t_0$. That is,

$$a_n(t_0) = \delta_{m,n}. \qquad (8.241)$$

Then,

$$\frac{da_k}{dt} = (1/\hbar\omega_{km})\langle u_k \left| \frac{\partial \hat{H}}{\partial t}\right| u_m\rangle \exp\{i\omega_{km}(t - t_0)\}, (k \neq m). \qquad (8.238^3)$$

Integrating from $t = t_0$ to $t = t$, we get,

$$a_k(t) = (1/\hbar\omega_{km})\langle u_k \left| \frac{\partial \hat{H}}{\partial t}\right| u_m\rangle \int_0^{(t-t_0)} \exp(i\omega_{km}t')dt'$$

$$= (i/\hbar\omega_{km}^2)\langle u_k \left| \frac{\partial \hat{H}}{\partial t}\right| u_m\rangle [1 - \exp\{i\omega_{km}(t - t_0)\}], \qquad (8.242)$$

41. See, Schiff, L.I., op.cit., Section 35.

so that

$$|a_k(t)|^2 \equiv w_k(t) = \frac{4}{\hbar^2 \omega_{km}^4} \left| \langle u_k | \frac{\partial \hat{H}}{\partial t} | u_m \rangle \right|^2 \cdot \sin^2 \left\{ \frac{1}{2} \omega_{km} (t - t_0) \right\}, \quad k \neq m. \tag{8.243}$$

This gives the probability for the state u_k, which does not evolve continuously from u_m, to be occupied at time t. The condition for the validity of the adiabatic approximation is that this probability be negligible. That is, $w(t) \ll 1$, or,

$$\left| \langle u_k | \frac{\partial \hat{H}}{\partial t} | u_m \rangle \right| \ll \hbar (\omega_{km})^2. \tag{8.244}$$

Since this condition should hold for all times t, where $t_0 \leq t \leq t_1$, we can use, in (8.244), the values of u_k, u_m and ω_{km} at the time t.

Adiabatic Reversal of a Magnetic Field

As an illustrative example of the application of the adiabatic approximation, we consider the effect of the slow (adiabatic) reversal of a magnetic field on the energy levels of an atom. We make the same assumptions regarding the nature of the atom and the time dependence and direction of the magnetic feld as in the case of the sudden approximation. Moreover, we assume that the atom is initially in the $2P$ state (that is, $L = 1$, $S = \frac{1}{2}$).

According to Eqs. (8.229) and (8.230), we have,

$$\hat{H}(t) = \hat{H}^{(0)} + \alpha(\hat{\mathbf{L}} \cdot \mathbf{S}) - \alpha\hbar\rho(t)(\hat{L}_z + 2\hat{S}_z), \tag{8.245}$$

where,

$$\rho(t) = \left(\frac{e\hbar}{2\mu c} \right) \frac{B(t)}{\alpha\hbar^2} = (\mu_B / \alpha\hbar^2) B(t), \tag{8.245a}$$

and

$$B(t) = B_0 \left\{ \frac{2(t - t_0)}{T} - 1 \right\}. \tag{8.245b}$$

The eigenvectors of $\hat{H}^{(0)}$ are $|LSM_L M_S>$, whereas the eigenvectors of $\hat{H}^{(0)} + \alpha(\hat{\mathbf{L}} \cdot \hat{\mathbf{S}})$ are $|LSJM_J>$, where $|L - S| \leq J \leq (L + S)$ and $M_J = M_L + M_S$ (that is, $\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}$). Thus, the $2P$ Level, which is 6-fold degenerate in the absence of the spin-orbit coupling, splits up into two levels, one of degeneracy 2 $\left(J = \frac{1}{2}\right)$ and the other of degeneracy 4 ($J = 3/2$) due to the spin-orbit coupling (see Fig. 8.13). But $(\hat{L}_z + 2\hat{S}_z)$ does not commute with $\hat{\mathbf{J}}^2$. As a result, when the magnetic field is strong enough to make the effect of the $(\hat{\mathbf{L}} \cdot \hat{\mathbf{S}})$ term negligible, the eigenvectors of $\hat{H}(t)$ would be approximately those of $\hat{H}^{(0)}$, namely $|LSM_L M_S>$. However, the degeneracy would be completely lifted (Fig. 8.13) as the eigenvalues of $\hat{H}(t)$ depend on the values of M_L and M_S (see Eq. (8.247) below). Thus, if we vary the magnetic field slowly from a large negative value to a large positive value, both the initial state and the final state would be characterised by definite values of M_L and M_S (and, hence, also by a definite value of $M_J = M_L + M_S$). But the transition

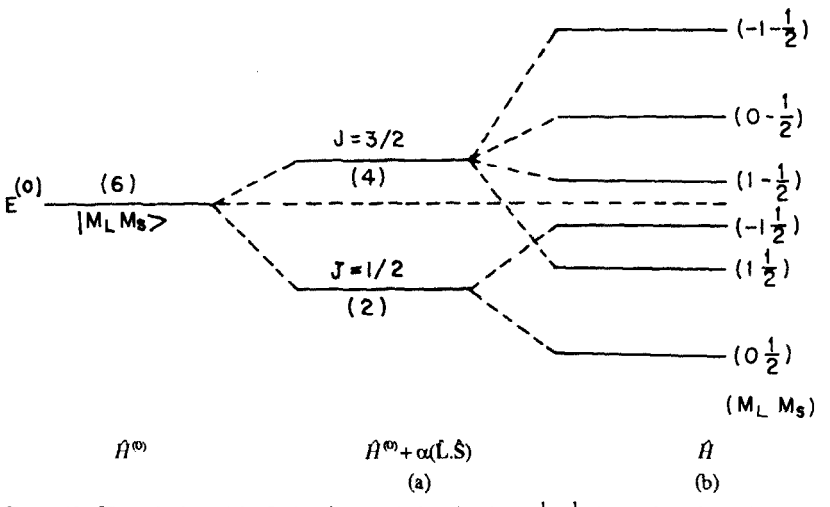


Fig. 8.13. The splitting of the $2p$ level ($M_L = 1, 0, -1$; $M_S = +\frac{1}{2}, -\frac{1}{2}$) due to (a) spin-orbit coupling, (b) magnetic field.

from the initial to the final state has to pass through a stage where the spin-orbit term dominates over the magnetic term. During this stage, the state of the system is characterised by a definite value of M_J but no definite individual values of M_S and M_L. This means that the system can make a transition from a state with one set of values for M_L and M_S to a state with another set of values for M_L and M_S such that the sum of M_L and M_S is preserved. Thus, if $|M_L M_S> \equiv |LSM_L M_S>$ is the initial state, $|M'_L M'_S> \equiv |LSM'_L M'_S>$ could be the state at the end of the reversal of the magnetic field, where $M'_L + M'_S = M_L + M_S$. The transition from $|M_L M_S>$ to $|M'_L M'_S>$ is not, however, a sudden jump but a continuous one described by the equation,

$$|M_J> = \sum_{M''_L M''_S} c_{M''_L M''_S} |M''_L M''_S>, (M''_L + M''_S = M_J), \qquad (8.246)$$

where the coefficients $c_{M''_L M''_S}$ are continuous functions of time such that

$$c_{M''_L M''_S}(t_0) = \delta_{M_L M''_L} \delta_{M_S M''_S}$$

and

$$c_{M''_L M''_S}(t_1) = \delta_{M'_L M''_L} \delta_{M'_S M''_S}.$$

The continuity of the state, here, is characterised by a constant value[42] for the eigenvalue of $\hat{J}_z$ throughout the period t_0 to t_1. The energies of the levels are also continuous functions of the parameter $\rho(t)$ [defined by Eq. (8.245a)]. The variation of the energies with ρ is shown in Fig. 8.14. This could be obtained using the relationship,

42. Of course, a constant value of M_J is only a necessary, and not a sufficient, condition for the continuity of the evolution.

$$\frac{1}{\alpha\hbar^2}< M'_L\, M'_S \mid \hat{H}(t) \mid M_L\, M_S >$$

$$= \{M_L\, M_S - (M_L + 2M_S)\rho(t)\}\,\delta_{M_L M'_L}\delta_{M_S M'_S}$$

$$+\frac{1}{2}\left\{\left(M_S+\frac{1}{2}\right)\left(\frac{3}{2}-M_S\right)(1-M_L)(2+M_L)\right\}^{1/2}$$

$$\times\,\delta_{M_L+1,M'_L}\delta_{M_S-1,M'_S}+\frac{1}{2}\left\{\left(\frac{1}{2}-M_S\right)\left(\frac{3}{2}+M_S\right)\right.$$

$$\left.\times(1+M_L)(2-M_L)\right\}^{1/2}\delta_{M_L-1,M'_L}\delta_{M_S+1,M'_S}, \qquad (8.247)$$

where, we have put $< M_L M_S \mid \hat{H}^{(0)} \mid M_L M_S > = 0$. Let us illustrate with two specific cases.

Case 1. $M_J = 3/2$: There is only one state with $M_J = 3/2$, namely $\mid M_L M_S\rangle = \mid 1\tfrac{1}{2}\rangle$. The adiabatic approximation, then, requires that the system remain in the state $\mid 1\tfrac{1}{2}\rangle$ throughout the reversal of the magnetic field if it is initially in that state. The energy of the state will, however, vary with time according to the formula [see Eq. (8.247)],

$$E\left(1\frac{1}{2}\right) = \langle 1\frac{1}{2} \mid \hat{H}(t) \mid 1\frac{1}{2}\rangle \qquad (8.248)$$

$$= \left(\frac{1}{2} - 2\rho(t)\right\}\alpha\hbar^2.$$

From Eq. (8.245), we have,

$$\frac{\partial \hat{H}}{\partial t} = -\alpha\hbar(\hat{L}_z + 2\hat{S}_z)\frac{d\rho}{dt}. \qquad (8.249)$$

Also, here, $\mid u_m\rangle \equiv \mid 1\tfrac{1}{2}\rangle;\ \mid u_k\rangle \equiv \mid M_L M_S\rangle \neq \mid 1\tfrac{1}{2}\rangle,$

so that,

$$\langle u_k\left|\frac{\partial \hat{H}}{\partial t}\right|u_m\rangle = \langle M_L M_S \mid (\hat{L}_z + 2\hat{S}_z) \mid 1\frac{1}{2}\rangle \cdot \left\{-\alpha\hbar\frac{d\rho}{dt}\right\}$$

$$= 0, \text{ for } M_L \neq 1;\ M_S \neq \frac{1}{2}.$$

Case 2. $M_J = \frac{1}{2}$: There are two states with $M_J = \frac{1}{2}$, namely, $\left(0\tfrac{1}{2}\right\rangle$ and $\mid 1\ -\tfrac{1}{2}\rangle$. If the system is in one of these states at $t = t_0$, then the wavefunction of the system at $t_0 < t < t_1$ would be given, according to Eq. (8.246), by

Thus, the condition (8.244) is very well satisfied.

$$\left|\frac{1}{2}\right\rangle_p = c^{(p)}_{0\frac{1}{2}}\left|0\frac{1}{2}\right\rangle + c^{(p)}_{1-\frac{1}{2}}\left|1-\frac{1}{2}\right\rangle,\ p=1,2, \tag{8.250}$$

where,

$$\left|c^{(p)}_{0\frac{1}{2}}\right|^2 + \left|c^{(p)}_{1-\frac{1}{2}}\right|^2 = 1.$$

The amplitudes $c^{(p)}_{M_L M_S}$ as well as the energies of the two states could be obtained by diagonalizing the Hamiltonian matrix corresponding to $M_J = \frac{1}{2}$. Using Eq. (8.247), we get,

$$H\left(M_J=\frac{1}{2}\right) = \alpha\hbar^2\begin{pmatrix} -\rho & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\frac{1}{2}\end{pmatrix} \tag{8.251^1}$$

The eigenvalues and eigenvectors of this matrix could be determined by the method described in Section A7. But, it turns out that the following procedure is simpler:

Any 2×2 matrix could be written as a linear sum of the Pauli spin matrices $\sigma_x, \sigma_y, \sigma_z$ given by (5.34), and the unit matrix $I = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$. Thus,

$$\begin{pmatrix} -\rho & \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} & -\frac{1}{2}\end{pmatrix} = b_0 I + b_x\sigma_x + b_y\sigma_y + b_z\sigma_z$$

$$= \begin{pmatrix} b_0+b_z & b_x - ib_y \\ b_x+ib_y & b_0-b_z\end{pmatrix}$$

Hence,
$$b_0 = -\frac{1}{4}(1+2\rho),$$

$$b_x = \frac{1}{\sqrt{2}};\ b_y = 0;\ b_z = \frac{1}{4}(1-2\rho), \text{ and}$$

$$H\left(M_J=\frac{1}{2}\right) = \alpha\hbar^2\left[-\frac{1}{4}(1+2\rho)I + (\mathbf{b}\cdot\sigma)\right], \tag{8.251^2}$$

where $\mathbf{b}$ is the vector whose components are b_x, b_y and b_z and whose magnitude is

$$b = (b_x^2 + b_y^2 + b_z^2)^{1/2} = \frac{1}{4}\sqrt{8+(1-2\rho)^2}. \tag{8.252}$$

The eigenvalues of $(\mathbf{b}\cdot\sigma)$ are $\pm b$. This follows from the fact that the component of σ along any direction has the values ± 1.

Problem 8.17: Verify that the eigenvalues of $(\mathbf{b}\cdot\sigma)$ are $\pm b$, from the secular equation of $(\mathbf{b}\cdot\sigma)$.

The eigenvalues of $\hat{H}$ are thus:

$$E_+(t) \equiv E_1(t) = \alpha\hbar^2 \epsilon_+, \tag{8.253a}$$

$$E_-(t) \equiv E_2(t) = \alpha\hbar^2 \epsilon_-, \tag{8.253b}$$

where,
$$\epsilon_\pm = -\frac{1}{4}(1+2\rho) \pm b. \tag{8.253}$$

The corresponding eigenvectors are:

$$|+\rangle \equiv |\frac{1}{2}\rangle_1 = c^{(+)}_{0\frac{1}{2}} |0\frac{1}{2}\rangle + c^{(+)}_{1-\frac{1}{2}} |1-\frac{1}{2}\rangle, \tag{8.250a}$$

$$|-\rangle \equiv |\frac{1}{2}\rangle_2 = c^{(-)}_{0\frac{1}{2}} |0\frac{1}{2}\rangle + c^{(-)}_{1-\frac{1}{2}} |1-\frac{1}{2}\rangle, \tag{8.250b}$$

with
$$c^{(\pm)}_{0\frac{1}{2}} = \frac{1}{[1+2(\rho+\epsilon_\pm)^2]^{1/2}} \tag{8.254a}$$

$$c^{(\pm)}_{1-\frac{1}{2}} = \frac{\sqrt{2}(\rho+\epsilon_\pm)}{[1+2(\rho+\epsilon_\pm)^2]^{1/2}}. \tag{8.254b}$$

At time $t = t_0$, ρ is large and negative [see Eqs. (8.245a, b)];

$$\rho = -\rho_0 = -(\mu_B/\alpha\hbar^2)B_0 \ll 1.$$

Then, from Eqs. (8.252), (8.253) and (8.254a, b) we have,

$$b \approx \frac{1}{4}(1+2\rho_0);$$

$$\epsilon_+ \approx +\rho_0 \gg 1, \epsilon_- \approx -\frac{1}{2};$$

$$c^{(+)}_{0\frac{1}{2}} \approx 1, c^{(+)}_{1-\frac{1}{2}} \approx 0; |+\rangle = |0\frac{1}{2}\rangle,$$

$$c^{(-)}_{0\frac{1}{2}} \approx 0, c^{(-)}_{1-\frac{1}{2}} \approx 1; |-\rangle = |1-\frac{1}{2}\rangle.$$

At time $t = t_1$, $\rho = \rho_0 \gg 1$.

$$b \approx 2\rho_0$$

$$\epsilon_+ \approx -\frac{1}{4}; \epsilon_- \approx -\rho_0.$$

$$c^{(+)}_{0\frac{1}{2}} \approx \frac{1}{\sqrt{2}\rho_0} \approx 0, c^{(+)}_{1-\frac{1}{2}} \approx 1; |+\rangle = |1-\frac{1}{2}\rangle$$

$$c^{(-)}_{0\frac{1}{2}} \approx 1; c^{(-)}_{1-\frac{1}{2}} \approx 0; |-\rangle = |0\tfrac{1}{2}\rangle.$$

Thus, as the magnetic field is varied from $-B_0$ to $+B_0$, the state of the system varies continuously from $|0\frac{1}{2}\rangle$ to $|1-\frac{1}{2}\rangle$ or from $|1-\frac{1}{2}\rangle$ to $|0\frac{1}{2}\rangle$. This variation (along with the variation in the energies) is shown in Fig. 8.14 where the $M_J = \frac{1}{2}$ levels are labelled $|+>$ and $|->$.

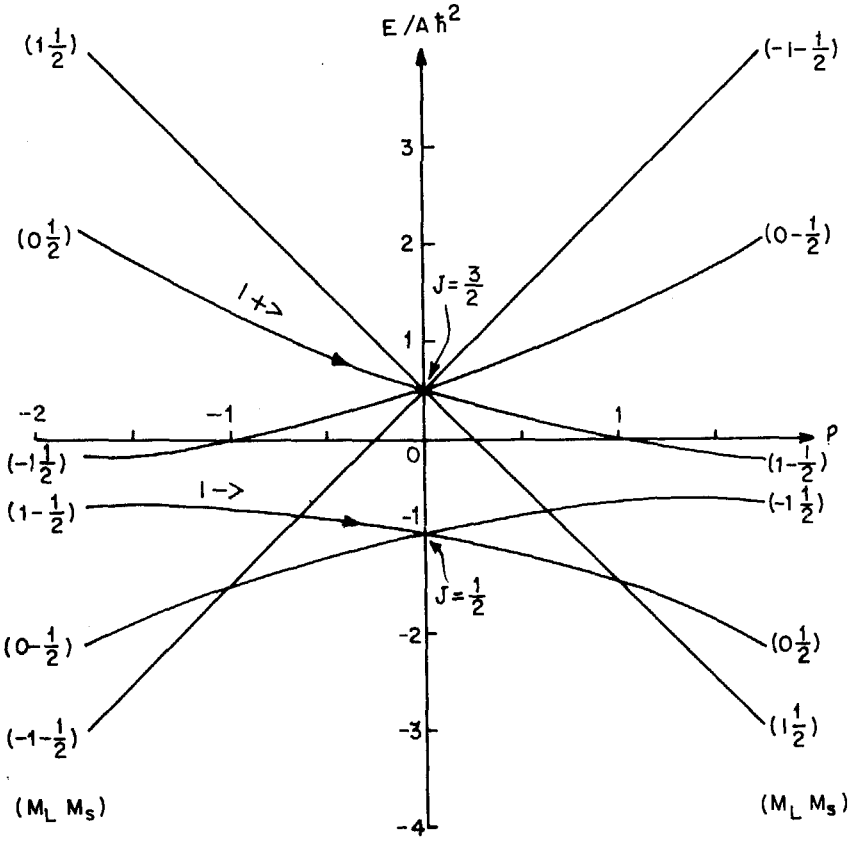


Fig. 8.14. The energies and the wave functions of the $2P$ levels as a function of the intensity of the magnetic field. The levels labelled $|+>$ and $|->$ are the two states with $M_J = \frac{1}{2}$. The numbers in parentheses are the quantum numbers $(M_L M_S)$ of the states in the two limits.

The condition for the validity of this adiabatic description (that is, for the system not to make a transition from $|+>$ to $|->$ or vice versa during the reversal of the field) is [according to (8.244)].

$$\langle - \left| \frac{\partial \hat{H}}{\partial t} \right| + \rangle \ll \hbar\omega_{+-}^2 = (E_+ - E_-)^2/\hbar = 4\alpha^2\hbar^3 b^2,$$

or,

$$\frac{\left|\langle - \left| \frac{\partial \hat{H}}{\partial t} \right| + \rangle\right|}{4\alpha^2\hbar^3 b^2} \ll 1, \qquad (8.255^1)$$

where $\frac{\partial \hat{H}}{\partial t}$ is given by Eq. (8.249). We have,

$$\langle - \left| \frac{\partial \hat{H}}{\partial t} \right| + \rangle = -\left(\hbar \frac{d\rho}{dt}\right)\alpha < - | (\hat{L}_z + 2\hat{S}_z) | + \rangle$$

$$= -\alpha\hbar^2 \frac{d\rho}{dt} \cdot c^{(+)}_{0\frac{1}{2}} c^{(-)}_{0\frac{1}{2}},$$

where, expressions (8.250a, b) for $|+>$ and $|->$ have been used. Since $c^{(+)}_{0\frac{1}{2}} c^{(-)}_{0\frac{1}{2}} \leq 1$, we get,

$$\left|\langle - \left| \frac{\partial \hat{H}}{\partial t} \right| + \rangle\right| \leq \alpha\hbar^2 \left| \frac{d\rho}{dt} \right|,$$

and condition (8.255^1) reduces to:

$$\frac{\alpha\hbar^2 \left| \frac{d\rho}{dt} \right|}{4\alpha^2\hbar^3 b^2} \ll 1. \qquad (8.255^2)$$

Now, the L.H.S. of (8.255^2) is maximum when $b^2 = \frac{1}{16}[8 + (1 - 2\rho)^2]$ is minimum. That is, when $\rho = b^2 = \frac{1}{2}$. Also, from Eqs. (8.245a, b) we have, $\left| \frac{d\rho}{dt} \right| = 2\mu_B B_0 / \alpha\hbar^2 T$. The condition for the validity of the adiabatic approximation, then, becomes,

$$\frac{2\mu_B B_0}{2\alpha^2\hbar^3 T} \ll 1,$$

or,

$$T' \equiv \left(\frac{2\alpha\hbar^2}{2\mu_B B_0}\right) T \gg \frac{1}{\alpha\hbar}. \qquad (8.255^3)$$

Now, T' is the time during which the magnetic energy changes from $-\alpha\hbar^2$ to $+\alpha\hbar^2$ (since during T, the magnetic energy changes from $-\mu_B B_0$ to $+\mu_B B_0$). It is during this time that the spin-orbit energy (which is of the order of $\alpha\hbar^2$) dominates over the magnetic energy, and it is also during this period that the system changes continuously from the state $|1 - \frac{1}{2}>$ to $|0\frac{1}{2}>$ or vice versa. On the other hand, $(1/\alpha\hbar)$ is the characteristic time associated with the spin-orbit splitting. That is, $(1/\alpha\hbar)$ is the time characterising a discontinuous (or discrete) transition $|+> \leftrightarrow |->$ taking place under the influence of the spin-orbit interaction. The

reversal of the magnetic field should take place very slowly compared with this time for the transition to be adiabatic. This condition may be contrasted with the condition (8.233) in the case of the sudden approximation.

Problem 8.18: Consider a linear harmonic oscillator whose equilibrium position x_0 depends linearly on the time. If the oscillator is initially in its ground state, obtain the conditions for the applicability of (a) the sudden approximation, (b) the adiabatic approximation.

CHAPTER 9

IDENTICAL PARTICLES

9.1 THE IDENTITY OF PARTICLES

In Chapter 6, we have discussed certain symmetries which arise from the properties of space and time. In this Chapter, we are concerned with a different type of symmetry which is encountered in the quantum mechanical behaviour of a system of identical particles. Now, the epithet *identical*, as applied to a group of particles, needs some clarification. We say all electrons are identical because one electron cannot be distinguished from another electron by means of any of its inherent physical attributes such as mass, electric charge, spin, etc. The fact that one electron might be in a spin-up state while the other is in a spin-down state, or that the electrons might be having different momenta, is not a hindrance to their being considered identical. This is because different values of the spin-component or linear momentum merely designate different dynamical *states* of an electron. Thus, what is required for two particles to be identical is that each of them should be described by the same complete, commuting set of observables, besides being identical in the physical attributes that are not described as eigenvalues of these observables. In the example of the electrons, the operators $\hat{p}_z$ and $\hat{s}_z$ representing respectively the components of the linear momentum and the spin, constitute a complete commuting set of observables.

The above criterion for the identity of particles has the following interesting consequence. Particles that are different from the viewpoint of their physical attributes such as mass and electric charge, could be considered identical if we could ascribe the differences to the different eigenvalues of one or more observables that commute with the Hamiltonian of the system. The case of the proton and the neutron is an example. The proton is positively charged while the neutron has no charge. Their masses are also different though the difference is small. They could be described as two different states of the same particle the *nucleon*, by introducing an observable called the *isospin* (analogous to the spin). For the nucleon, the isospin $t = \frac{1}{2}$ so that the proton and the neutron correspond to the two possible values of t_z, namely $t_z = \pm\frac{1}{2}$ The description of the proton and the neutron as different states of the same particle is, however, not valid when electromagnetic interaction is involved since the Hamiltonian in that case does not commute with the isospin.

An important aspect of the symmetry associated with a system of identical particles is that it is intimately related to the spin of the particles. As a result, we will not be able to ignore the spin as we have done so far in most of our discussions.

The Indistinguishability Principle

In classical mechanics, identical particles can, in principle, be distinguished from one another either by observing their individual paths or by means such as labelling them. This is so because observation or labelling does not affect the classical dynamics (and hence the classical paths) of the particles. In the case of a system of identical micro-particles, on the other hand, observation of the individual particles is not possible without seriously affecting the dynamics of the system. For example, in the scattering of protons by protons, the proton which comes to the detector could be either the projectile or the target proton, but it would be impossible to tell them apart. Similarly, in a Helium atom a state in which one of the electrons is in the quantum state ϕ_α and the other electron is in the state ϕ_β would be indistinguishable from a state in which the electrons are interchanged. This inability to tell apart the particles from one another in a system of identical particles is embodied in the *principle of indistinguishability of identical particles.* Quantum mechanics of a system of identical particles should be so formulated as to be consistent with this principle. For example, in the scattering problem mentioned above, experiment does not distinguish between the projectile and the target. Therefore, in the quantum mechanical calculation of the scattering cross section also, no distinction should be permitted between the projectile and the target.

It is to be emphasized that the significance of the principle of indistinguishability is far deeper in quantum mechanics than it is in classical mechanics. The peculiar laws underlying quantum mechanics, in particular the principle of superposition, are responsible for this. For instance, the indistinguishability between the projectile and the target protons in the scattering of protons implies only the following in classical mechanics: The observed cross section would be the sum of the cross sections corresponding to the cases (a) and (b) in Fig. 9.1. But in the case of quantum mechanics, it is the scattering amplitudes corresponding to the two cases that are to be added (or subtracted, depending on the spin of the particles) to obtain the scattering amplitude corresponding to the observed cross section. Since the scattering cross section is the absolute square of the scattering amplitude, there is no simple relationship between the quantum mechanical cross section and the cross sections corresponding separately to cases (a) and (b) in Fig. 9.1.

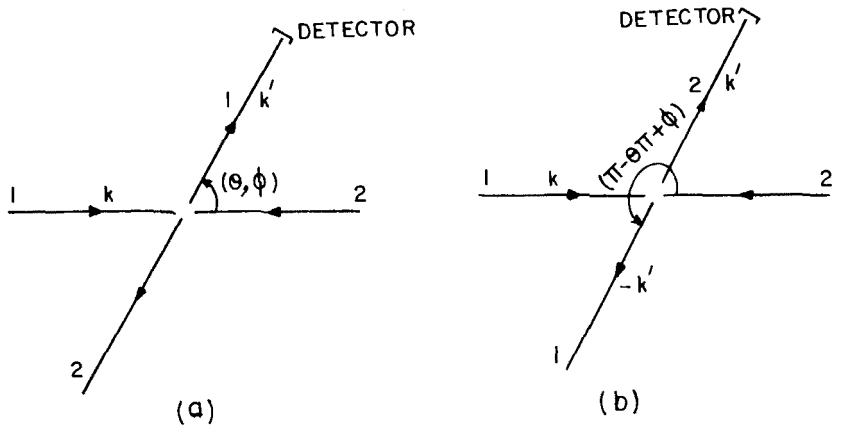


Fig. 9.1. Scattering of identical particles in the Centre-of-Mass system.

Symmetry of Wavefunctions

In order to illustrate how the principle of indistinguishability influences the quantum mechanics of a system of identical particles, let us consider the simple case of a two-particle system. We use the following notations: The numbers inside the parenthesis following the symbol representing an operator or a wavefunction, label the particles whereas the position inside the parenthesis denotes the quantum mechanical states of the particles. Thus, if $\hat{H}(1,2)$ stands for the Hamiltonian of the system with particle number 1 in the state α and particle number 2 in state β, then $\hat{H}(2,1)$ represents the Hamiltonian when particle number 2 is in state α and particle number 1 is in state β.

Now, the principle of indistinguishability requires that the physical observables of the system, and in particular the Hamiltonian, be invariant under the interchange of the two particles. That is,

$$\hat{H}(1,2) = \hat{H}(2,1). \tag{9.1}$$

The Schrödinger equation of the system is,

$$i\hbar \frac{\partial \psi(1,2)}{\partial t} = \hat{H}(1,2)\psi(1,2), \tag{9.2}$$

where particle number 1 is in the state α and particle number 2 is in the state β. If we interchange the particles, the corresponding Schrödinger equation becomes [in view of (9.1)],

$$i\hbar \frac{\partial \psi(1,2)}{\partial t} = \hat{H}(1,2)\psi(2,1), \tag{9.3}$$

Eqs. (9.2) and (9.3) show that if $\psi(1,2)$ is a possible state of the two-particle system with a certain energy, then $\psi(2,1)$ is also a possible state with the *same* energy. From the viewpoint of classical mechanics, this is a trivial result. But the principle of superposition makes it a result of profound significance in quantum mechanics. According to this principle, if $\psi(1,2)$ and $\psi(2,1)$ are possible

solutions of the Schrödinger equation corresponding to a particular energy, then any linear superposition of $\psi(1,2)$ and $\psi(2,1)$ is also a possible solution belonging to the same energy eigenvalue. In particular, the linear combinations[1],

$$\psi_S(1,2) = \frac{1}{\sqrt{2(1+\delta_{\alpha\beta})}}[\psi(1,2) + \psi(2,1)], \tag{9.4a}$$

and

$$\psi_A(1,2) = \frac{1}{\sqrt{2}}[\psi(1,2) - \psi(2,1)], \tag{9.4b}$$

are solutions of the Schrödinger equation for the system. These linear combinations differ from $\psi(1,2)$ and $\psi(2,1)$ in that they have definite symmetry properties under the interchange of the particles.
For,

$$\psi_S(2,1) = \psi_S(1,2); \quad \psi_A(2,1) = -\psi_A(1,2). \tag{9.5}$$

We say that ψ_S is *symmetric* while ψ_A is *antisymmetric*. The importance of ψ_S and ψ_A arises from the fact that these are the *only* wave functions that are consistent with the principle of indistinguishability of identical particles. This follows from the mathematical structure of quantum mechanics, as shown below:

Let $\hat{P}_{12}$ represent the operator that exchanges particles 1 and 2. That is,

$$\hat{P}_{12}\psi(1,2) = \psi(2,1). \tag{9.6}$$

Obviously,

$$\hat{P}_{12}^2 = \hat{1}. \tag{9.7}$$

Thus, $\hat{P}_{12}$ is an involution (see *Projection Operators*, Section 2.2) and is hence both *Hermitian* and *Unitary* like the parity operator (Section 6.2D). Like the parity operator too, $\hat{P}_{12}$ represents a discrete transformation. The invariance of the Hamiltonian under this transformation (that is required by the principle of indistinguishability) is expressed by the condition [see Eq. (6.3a)],

$$[\hat{P}_{12}, \hat{H}] = \hat{0}. \tag{9.8[1]}$$

This means that only those eigenvectors of $\hat{H}$ are permissible that are also simultaneously eigenvectors of the *exchange operator* $\hat{P}_{12}$. But according to Eq. (9.7), the eigenvalues of $\hat{P}_{12}$ are +1 and −1. Denoting the corresponding eigenvectors by $\psi_+(1,2)$ and $\psi_-(1,2)$, we have

$$\hat{P}_{12}\psi_+(1,2) \equiv \psi_+(2,1) = +\psi_+(1,2), \tag{9.9a}$$

$$\hat{P}_{12}\psi_-(1,2) \equiv \psi_-(2,1) = -\psi_-(1,2), \tag{9.9b}$$

Comparing these with (9.5), we see that

---

1. The factors $1/\sqrt{2}$ and $1/\sqrt{2(1+\delta_{\alpha\beta})}$ in Eqs. (9.4a, b) are normalizing factors.

$$\psi_+ = \psi_S(1,2) = \{1/\sqrt{2(1+\delta_{\alpha\beta})}\}\,(\hat{1}+\hat{P}_{12})\psi(1,2), \tag{9.10a}$$

$$\psi_- = \psi_A(1,2) = (1/\sqrt{2})\,(\hat{1}-\hat{P}_{12})\psi(1,2). \tag{9.10b}$$

Thus, the physical state of a system of two identical particles is represented by either ψ_S or ψ_A. Moreover, Eq. (9.8^1) implies that the symmetry is a constant of motion [cf. Eq. (6.1)]. Therefore, if the system is, say, in a symmetric state at some time, then it will continue to be in the symmetric state for all time.

The foregoing considerations are easily extended to the case of a system of n identical particles. In this case, we have $\binom{n}{2} = (n/2)\,(n-1)$ two-particle exchange operators $\hat{P}_{ij}$ all of which should commute with the Hamiltonian of the system:

$$[\hat{P}_{ij}, \hat{H}] = \hat{0}., (i = 1,2,\ldots n-1;\ j = 2,3,\ldots,n). \tag{9.8^2}$$

However, the exchange operators do not commute among themselves:

$$\hat{P}_{ij}\hat{P}_{jk} \neq \hat{P}_{jk}\hat{P}_{ij}. \tag{9.10}$$

As a result, it is not possible to find a *complete set* of functions that are simultaneous eigenvectors of all the exchange operators and $\hat{H}$ (except for the case $n = 2$ where there is only one exchange operator). But the principle of indistinguishability expressed by Eq. (9.8^2) requires the existence of at least one wavefunction that is a simultaneous eigenvector of all the exchange operators and $\hat{H}$. Let Φ denote such a function. Then,

$$\hat{H}\Phi = E\Phi, \tag{9.11a}$$

$$\hat{P}_{ij}\Phi = \lambda_{ij}\Phi, (i = 1,2,\ldots,n-1;\ j = 2,3,\ldots,n). \tag{9.11b}$$

where

$$\lambda_{ij} = \pm 1.$$

Now,

$$\hat{P}_{ij}\hat{P}_{ik} = \hat{P}_{jk}\hat{P}_{ij} = \hat{P}_{ik}\hat{P}_{jk}, \tag{9.12}$$

for any three different values of i, j, k.

Eqs. (9.12) and (9.11b) give,

$$\lambda_{ij}\lambda_{ik} = \lambda_{jk}\lambda_{ij} = \lambda_{ik}\lambda_{jk},$$

or,

$$\lambda_{ij} = \lambda_{ik} = \lambda_{jk}. \tag{9.13}$$

That is, the common eigenvector Φ belongs to the *same* eigenvalue of all the exchange operators. If $\lambda_{ij} = +1$, we say that Φ is *totally symmetric* (it is symmetric under the interchange of *any* pair of particles) and denote it by Φ_S whereas if $\lambda_{ij} = -1$, Φ is *totally antisymmetric* and is denoted by Φ_A:

$$\hat{P}_{ij}\Phi_S = +\Phi_S;\ \hat{P}_{ij}\Phi_A = -\Phi_A. \tag{9.14}$$

Thus, we have the important result: *The wavefunction of a system of identical particles is either totally symmetric or totally antisymmetric.*

Problem 9.1: Verify relationship (9.12) for $n = 3$.

9.2 SPINS AND STATISTICS

We have seen that the symmetry character of a system of identical particles is a constant of motion. It is also found that a given type of particle is associated with only one type of symmetry. Thus a system of electrons is always described by an antisymmetric wavefunction while a system of pions is invariably described by a symmetric wavefunction. That is, the symmetry character of the wavefunction is an intrinsic property of the particles.

Now, the statistical properties of a system of a large number of particles depend on the available degrees of freedom for each value of the total energy of the system. For a system of distinguishable particles, every permutation of the particles gives rise to a different state. Such a system obeys *classical* or *Boltzmann* statistics. The statistics obeyed by a system of identical particles with symmetric wave functions is known as *Bose-Einstein* statistics while identical particles with anti-symmetric wavefunctions obey *Fermi-Dirac* statistics. Since the symmetry character of the wavefunction is an intrinsic property of the particles, it follows that the statistics associated with a particle is also of an intrinsic nature. This fact justifies the classification of particles on the basis of the statistics obeyed by them. Thus, particles obeying Bose-Einstein statistics are called *bosons* while those obeying Fermi-Dirac statistics are referred to as *fermions*.

It is further found, and shown plausible in quantum field theory,[2] that the statistics is intimately connected with the spin of the particles. Bosons have integral (including zero) spin whereas particles with half-integral spin are fermions. This correlation between spin and statistics applies not only to elementary particles but to composite particles (such as atoms and nuclei) as well. This is understandable since a composite particle composed of fermions will have integral or half-integral spin according as the number of fermions is even or odd. In the former case, an interchange between two such particles is equivalent to an even number of interchanges of fermions so that the wavefunction should be symmetric. Similarly, in the latter case the wavefunction should be antisymmetric.

The Pauli Exclusion Principle

We have seen that the wavefunction of a system of spin-half particles is antisymmetric. Consider a system of two spin-half particles. If Φ_{α_1} and Φ_{α_2} denote the two quantum states available to the particles, then, $\psi(1, 2) = \Phi_{\alpha_1}(1)\Phi_{\alpha_2}(2)$, so that the wavefunction of the system is given by [see Eq. (9.10b)],

---

2. Pauli, W., *Phys. Rev.*, **58**, 716 (1940). Also, see Section 11.4B

$$\psi_A(1,2) = \frac{1}{\sqrt{2!}}(\hat{1} - \hat{P}_{12})\Phi_{\alpha_1}(1)\Phi_{\alpha_2}(2)$$

$$= \frac{1}{\sqrt{2!}}\begin{vmatrix} \Phi_{\alpha_1}(1) & \Phi_{\alpha_1}(2) \\ \Phi_{\alpha_2}(1) & \Phi_{\alpha_2}(2) \end{vmatrix}. \tag{9.15}$$

Obviously, $\psi_A(1,2)$ vanishes when $\Phi_{\alpha_1} = \Phi_{\alpha_2}$. That is, *it is not possible for the two particles to occupy the same state.* This is known as the *Pauli exclusion principle.*

Eq. (9.15) is easily generalized to the case of n particles. In this case, the wavefunction will be a linear superposition of the $n!$ functions that correspond to the $n!$ permutations of the particles:

$$\psi_A(1,2,\ldots,n) = \frac{1}{\sqrt{n!}}\sum_P (-1)^{\pi(P)}\hat{P}\Phi_{\alpha_1}(1)\Phi_{\alpha_2}(2)\ldots\Phi_{\alpha_n}(n)$$

$$= \frac{1}{\sqrt{n!}}\begin{vmatrix} \Phi_{\alpha_1}(1) & \Phi_{\alpha_1}(2) & \ldots & \Phi_{\alpha_1}(n) \\ \Phi_{\alpha_2}(n) & \Phi_{\alpha_2}(2) & \ldots & \Phi_{\alpha_2}(n) \\ \Phi_{\alpha_n}(1) & \Phi_{\alpha_n}(2) & \ldots & \Phi_{\alpha_n}(n) \end{vmatrix} \tag{9.16}$$

Here $\hat{P}$ represents one of the $n!$ permutations and $\pi(P)$ = the number of two particle exchange operators contained in $\hat{P}$. The determinant in (9.16) is known as the *Slater determinant.* Since the determinant vanishes when any two rows are identical (Section A3), the exclusion principle follows.

For a system of bosons, on the other hand, the wavefunction is given by,

$$\psi_S(1,2,\ldots,n) = \frac{1}{\sqrt{n!\delta}}\sum_P \hat{P}\Phi_{\alpha_1}(1)\Phi_{\alpha_2}(2)\ldots\Phi_{\alpha_n}(n); \tag{9.17}$$

where δ is a factor which depends on the number of particles occupying the same state. (9.17) does not vanish when any two α_i are equal. Thus, there is no exclusion principle for bosons.

9.3 ILLUSTRATIVE EXAMPLES

The following examples are chosen so as to illustrate the important role played by the symmetry of the wavefunction in the dynamics of both bound and unbound systems of identical particles.

The Helium Atom

From the viewpoint of atomic properties, the He atom is a system of two electrons. Since electrons are fermions, the total wavefunction of the system must be antisymmetric. However, the total wavefunction is the product of a space or orbital part and a spin part. Therefore, the space part alone could be symmetric or antisymmetric depending on whether the spin part is antisymmetric or symmetric. Since electron has spin $\frac{1}{2}$, the spin of a two-electron system could be either 1 or 0

[see Eq. (5.73b)]. There are three states with spin 1 corresponding to the three values (1, 0 and – 1) of the z-component of the spin. The spin-1 states are, hence, called the *triplet states*. We denote these by[3] ${}^3\chi_1$, ${}^3\chi_0$ and ${}^3\chi_{-1}$. Similarly, ${}^1\chi_0$ represents the *singlet* state corresponding to spin 0. From Eqs. (5.68a) and (5.83b), it follows that the triplet states are symmetric while the singlet state is antisymmetric. Thus, out of the eight combinations possible with the four spin functions and the two (the symmetric and the antisymmetric) space functions, only four are permitted by the principle of indistinguishability of identical particles. These are:

$$\Psi_S{}^1\chi_0,\ \Psi_A{}^3\chi_1,\ \Psi_A{}^3\chi_0,\ \Psi_A{}^3\chi_{-1}, \tag{9.18}$$

where [cf. Eqs. (9.10a, b)],

$$\Psi_S(1,2) = [1/2(1+\delta_{\alpha\beta})]^{1/2}(\hat{1}+\hat{P}_{12})\Phi_\alpha(1)\Phi_\beta(2), \tag{9.19a}$$

$$\Psi_A(1,2) = [1/\sqrt{2}]\,(\hat{1}-\hat{P}_{12})\Phi_\alpha(1)\Phi_\beta(2). \tag{9.19b}$$

Now, the Hamiltonian of the He atom can be written as

$$\hat{H} = \hat{H}_0 + \hat{H}_1, \tag{9.20}$$

where

$$\hat{H}_0 = \frac{-\hbar^2}{2\mu}(\nabla_1^2+\nabla_2^2) - 2e^2\left(\frac{1}{r_1}+\frac{1}{r_2}\right). \tag{9.20a}$$

$$\hat{H}_1 = \frac{e^2}{r_{12}}. \tag{9.20b}$$

Here, r_1 and r_2 are the radial distances of the electrons from the centre of the atom and r_{12} is the separation between the electrons. $\hat{H}_1$ represents the mutual repulsion of the electrons and is small in comparison with $\hat{H}_0$. The contribution of $\hat{H}_1$ to the energy of the system can, therefore, be calculated using first order perturbation theory (Section (8.3a)) where the zero-order wavefunctions (the eigenvectors of $\hat{H}_0$) are given by (9.18). Since $\hat{H}_1$ does not contain the spin variables, the three triplet states would be degenerate. And the separation between the singlet and the triplet states arises entirely due to the different symmetry of the orbital part.

According to Eq. (8.124b) the contribution to the energy due to $\hat{H}_1$, in the first order, is equal to the expectation value of $\hat{H}_1$. Thus, for the singlet state,

$$\begin{aligned} E_s^{(1)} &= \langle \Psi_S{}^1\chi_0 \mid \hat{H}_1 \mid \Psi_S{}^1\chi_0 \rangle \\ &= \langle \Psi_S \mid \hat{H}_1 \mid \Psi_S \rangle \\ &= \{1/(1+\delta_{\alpha\beta})\}\,(J_{\alpha\beta}+K_{\alpha\beta}), \end{aligned} \tag{9.21}$$

with

$$J_{\alpha\beta} = \langle \Phi_\alpha(1)\Phi_\beta(2) \mid \hat{H}_1 \mid \Phi_\alpha(1)\Phi_\beta(2) \rangle, \tag{9.21a}$$

---

3. The notation is ${}^{(2S+1)}\chi_{S_z}$ where S is the spin and S_z its projection along the z-axis.

$$K_{\alpha\beta} = \langle \Phi_\alpha(1)\Phi_\beta(2) \mid \hat{H}_1 \mid \Phi_\beta(1)\Phi_\alpha(2) \rangle, \tag{9.21b}$$

where expression (9.19a) for Ψ_S and property (9.8[1]) of $\hat{P}_{12}$ are used.

Similarly, for the triplet states,

$$E_t^{(1)} = \langle \Psi_A \mid \hat{H}_1 \mid \Psi_A \rangle = (J_{\alpha\beta} - K_{\alpha\beta}). \tag{9.22}$$

$J_{\alpha\beta}$ and $K_{\alpha\beta}$ are positive since $\hat{H}_1$ is positive definite. Therefore,

$$E_t^{(1)} < E_s^{(1)}.$$

That is, the singlet state is always higher in energy than the triplet state.

For the ground state of He atom, $\alpha = \beta \equiv 1S$ (that is, $n = 1, l = 0$), so that $\Psi_A \equiv 0$. Hence the only allowed state is $\Psi_S{}^1\chi_0$. Then, [see Eq. (8.90)],

$$E^{(1)} \equiv E_s^{(1)} = J_{(1s)^2} = \frac{5e^2}{4a_0} = 33.99 \text{ eV};$$

where a_0 is given by Eq. (8.85). Also, $e^2/a_0 = 27.19$ eV. From Eqs. (8.87a, b), we have,

$$E_{(1s)^2}^{(0)} = -4(e^2/a_0) = -108.76 \text{ eV},$$

so that,

$$E_{(1s)^2} = E_{(1s)^2}^{(0)} + E_{(1s)^2}^{(1)} = -74.77 \text{ eV}.$$

Similarly, we obtain the energies of some of the excited states, using the values,

$$J_{1s,2s} = 11.42 \text{ eV};$$

$$K_{1s,2s} = 1.20 \text{ eV};$$

$$J_{1s,2p} = 13.22 \text{ eV};$$

$$K_{1s,2p} = 0.94 \text{ eV}.$$

The energies so obtained are compared with the experimental values in Table 9.1. The spectroscopic notation ${}^{(2S+1)}L_J$ is used to specify a state, where S represents the spin and J the total angular momentum while L stands for the symbol that denotes the value of the orbital angular momentum according to the following scheme:

$$\text{Symbol for } L \rightarrow S \; P \; D \; F \ldots$$

$$\text{Value of } L \rightarrow 0 \; 1 \; 2 \; 3 \ldots \tag{9.23}$$

Thus, the symmetry character of the wavefunction, arising from the identity of the electrons, provides an understanding of the qualitative features (such as the nature of the ground state and the difference in energy between the triplet and the singlet states belonging to a given configuration) of the low lying levels of the He atom.

Table 9.1. Low-lying levels of the He atom.

Configuration (α, β)	State	Energy (eV)		ΔE^* $= E_s^{(1)} - E_t^{(1)}$
		Calculated	Experimental	(eV)
$(1s)^2$	$^1S_0$	−74.77	−78.98	—
$(1s, 2s)$	$^3S_1$	−57.77	−59.17	0.80 (2.39)
"	$^1S_0$	−55.38	−58.37	
$(1s, 2p)$	$^3P_{0,1,2}$	−55.71	−58.02	0.25 (1.85)
"	$^1P_1$	−53.86	−57.77	

*The numbers in the parenthesis represent the calculated values.

Problem 9.2: The Deuteron is a system of two nucleons, where the nucleon is a spin-half and isospin-half particle. In the ground state of the deuteron, both the nucleons occupy the $1s$ ($l = 0$) orbital state. Obtain the possible combinations of the total spin and the total isospin for the ground state.

Scattering of Identical Particles

In the scattering of identical particles, there are two situations indistinguishable from each other as shown in Fig. 9.1. The incident particle and the target have equal and opposite velocities in the Centre-of-Mass system. In (a) the particle observed at the detector is the incident particle (particle number 1) while in (b) it is the recoiled target (particle number 2). The indistinguishability of the two situations is taken into account by using properly symmetrised wavefunctions for the calculation of the scattering cross-section.

Now, for process (a), the asymptotic scattered wavefunction is [see Eq. (7.7)],

$$\psi_a^{(s)}(\mathbf{k}, \mathbf{k}', \mathbf{r}) \sim f_a(\mathbf{k}, \mathbf{k}') \frac{e^{ikr}}{r}. \tag{9.24}$$

Process (b) differs from process (a) in that particles 1 and 2 get interchanged at the time of scattering. Thus,

$$\psi_b^{(s)}(\mathbf{k}, \mathbf{k}', \mathbf{r}) = \hat{P}_{12} \psi_a^{(s)}(\mathbf{k}, \mathbf{k}', \mathbf{r}). \tag{9.25}$$

But interchanging the two particles is equivalent to reversing the relative coordinate $\mathbf{r} = (\mathbf{r}_1 - \mathbf{r}_2)$ and since, by definition, $\mathbf{k}'$ has the same direction as $\mathbf{r}$, it is equivalent to reversing the direction of $\mathbf{k}'$. Thus,

$$\hat{P}_{12} \psi_a^{(s)}(\mathbf{k}, \mathbf{k}', \mathbf{r}) = \psi_a^{(s)}(\mathbf{k}, -\mathbf{k}', \mathbf{r}). \tag{9.26}$$

A properly symmetrised wavefunction is a symmetric or an antisymmetric combination of $\psi_a^{(s)}$ and $\psi_b^{(s)}$:

$$\Psi_S^{(s)}(\mathbf{k},\mathbf{k}',\mathbf{r}) = \psi_a^{(s)}(\mathbf{k},\mathbf{k}',\mathbf{r}) + \psi_b^{(s)}(\mathbf{k},\mathbf{k}',\mathbf{r}), \tag{9.27^1}$$

$$\Psi_A^{(s)}(\mathbf{k},\mathbf{k}',\mathbf{r}) = \psi_a^{(s)}(\mathbf{k},\mathbf{k}',\mathbf{r}) - \psi_b^{(s)}(\mathbf{k},\mathbf{k}',\mathbf{r}). \tag{9.28^1}$$

Substituting from Eqs. (9.25), (9.26) and (9.24) and dropping the labels a, we have,

$$\Psi_S^{(s)} \sim f_S(\mathbf{k},\mathbf{k}')\,(e^{ikr}/r), \tag{9.27^2}$$

$$\Psi_A^{(s)} \sim f_A(\mathbf{k},\mathbf{k}')\,(e^{ikr}/r), \tag{9.28^2}$$

where,
$$f_S(\mathbf{k},\mathbf{k}') = f(\mathbf{k},\mathbf{k}') + f(\mathbf{k},-\mathbf{k}'), \tag{9.29}$$

$$f_A(\mathbf{k},\mathbf{k}') = f(\mathbf{k},\mathbf{k}') - f(\mathbf{k},-\mathbf{k}'). \tag{9.30}$$

The differential cross-section for the scattering of identical particles are then given by [Eq. (7.9)],

$$\sigma_S(\mathbf{k},\mathbf{k}') = |f_S(\mathbf{k},\mathbf{k}')|^2$$

$$= |f(\mathbf{k},\mathbf{k}')|^2 + |f(\mathbf{k},-\mathbf{k}')|^2 + 2\,\mathrm{Re}\,\{f^*(\mathbf{k},\mathbf{k}')f(\mathbf{k},-\mathbf{k}')\}, \tag{9.31^1}$$

and $\sigma_A(\mathbf{k},\mathbf{k}') = |f_A(\mathbf{k},\mathbf{k}')|^2$

$$= |f(\mathbf{k},\mathbf{k}')|^2 + |f(\mathbf{k},-\mathbf{k}')|^2 - 2\,\mathrm{Re}\,[f^*(\mathbf{k},\mathbf{k}')\cdot f(\mathbf{k},-\mathbf{k}')]. \tag{9.32^1}$$

These are to be compared with the corresponding classical expression,

$$\sigma_{cl}(\mathbf{k},\mathbf{k}') = \sigma(\mathbf{k},\mathbf{k}') + \sigma(\mathbf{k},-\mathbf{k}')$$

$$= |f(\mathbf{k},\mathbf{k}')|^2 + |f(\mathbf{k},-\mathbf{k}')|^2. \tag{9.33^1}$$

In (9.31^1) and (9.32^1), Re { } represents the real part of the quantity within the bracket. We see that an important feature that distinguishes the quantum mechanical cross sections from the classical one, is the *interference* between the scattering processes (a) and (b).

In terms of the angles θ and ϕ that specify the direction of scattering with respect to the incident direction,

$$f(\mathbf{k},\mathbf{k}') \equiv f_k(\theta,\phi),$$

and
$$f(\mathbf{k},-\mathbf{k}') \equiv f_k(\pi-\theta,\pi+\phi). \tag{9.34}$$

Further, in the case of *central forces*, the scattering amplitude is independent of the angle ϕ. Then,

$$f(\mathbf{k},\mathbf{k}') = f_k(\theta);\ f(\mathbf{k},-\mathbf{k}') = f_k(\pi-\theta),$$

and we have,

$$\sigma_S(\theta) = |f_k(\theta) + f_k(\pi-\theta)|^2, \tag{9.31^2}$$

$$\sigma_A(\theta) = |f_k(\theta) - f_k(\pi-\theta)|^2, \tag{9.32^2}$$

$$\sigma_{cl}(\theta) = |f_k(\theta)|^2 + |f_k(\pi-\theta)|^2, \tag{9.33^2}$$

Case of Spin $\frac{1}{2}$ Particles

The spin part of the wavefunction could be a triplet or a singlet state depending on whether the spins of the two particles are parallel or anti-parallel. The singlet state should be associated with the symmetric space function so that the corresponding cross-section is given by

$$^1\sigma_{1/2} = \sigma_S(\theta). \tag{9.35a}$$

Similarly, the triplet state is associated with the antisymmetric space function. Therefore,

$$^3\sigma_{1/2} = \sigma_A(\theta). \tag{9.35b}$$

The observed cross-section would be the weighted average of these two cases. That is,

$$\sigma_{1/2}(\theta) = \frac{1}{4}\sigma_S(\theta) + \frac{3}{4}\sigma_A(\theta)$$

$$= |f_k(\theta)|^2 + |f_k(\pi-\theta)|^2 - \text{Re}\{f_k^*(\theta) f_k(\pi-\theta)\}. \tag{9.36^1}$$

For $\theta = \pi/2$, this gives,

$$\sigma_{1/2}(\pi/2) = |f_k(\pi/2)|^2. \tag{9.36^2}$$

Case of Spin Zero Particles

In this case the total wavefunction is symmetric so that,

$$\sigma_0(\theta) = \sigma_S(\theta), \tag{9.37^1}$$

and

$$\sigma_0(\pi/2) = \sigma_S(\pi/2) = 4|f_k(\pi/2)|^2. \tag{9.37^2}$$

Also, from Eq. (9.33^2),

$$\sigma_{cl}(\pi/2) = 2|f_k(\pi/2)|^2. \tag{9.33^3}$$

Thus, we have the result,

$$\sigma_0(\pi/2):\sigma_{cl}(\pi/2):\sigma_{1/2}(\pi/2) = 4:2:1. \tag{9.38}$$

This result can be utilized to determine experimentally the spin of a particle like the proton. In this case, the classical cross section is given by the Rutherford formula (7.116). The fact that the observed cross section at $\theta = \pi/2$ is approximately half that of the value given by the Rutherford formula, indicates that the proton is a spin $\frac{1}{2}$ particle.

Problem 9.3: Show that the total cross sections for zero-energy scattering of identical particles of spin-zero and spin one-half are given, respectively, by

$$\sigma_0 = 16\pi a^2 \text{ and } \sigma_{1/2} = 4\pi a^2,$$

where 'a' is the scattering length.

REFERENCES

1. Powell, J.L. and Crasemann, B., *Quantum Mechanics* (Addison-Wesley, Massachusetts, 1961), Chapter 12.
2. Schiff, L.I., *Quantum Mechanics* (McGraw-Hill, 1968), 3rd edition, Chapter 10.

CHAPTER 10

RELATIVISTIC WAVE EQUATIONS

10.1 INTRODUCTION

The Schrödinger equation[1], which forms the basis of the mechanics we have considered so far, does not satisfy the requirements of the special theory of relativity, namely, invariance under Lorentz transformations[2]. This limitation restricts the applicability of the theory to systems with velocities small compared with the velocity of light. The first equation aimed at remedying this defect was obtained by Schrödinger himself[3]. But he discarded it because of, among other things, its failure to yield the correct spectrum of the hydrogen atom. The same equation was later proposed and discussed independently by Klein, Fock and Gordon[4]. It is commonly referred to as the *Klein-Gordon equation.*

The Klein-Gordon equation is derived on the basis of the following arguments: The 4-dimensional (Minkowski space) formalism of relativistic mechanics suggests the generalisation,

$$\hat{p}_\mu = -i\hbar \frac{\partial}{\partial x_\mu}, (\mu = 1, 2, 3, 4), \tag{10.1}$$

of Eq. (3.18[1]), where p_μ is a component of the *four-momentum* p. The space components (that is, the first three components) of p are the components of the momentum vector[5] $\mathbf{p}$ while,

$$p_4 = (i/c)E, \tag{10.2}$$

---

1. The reference here is to Eq. (4.14) with the Hamiltonian $\hat{H}$ derived, in accordance with Postulate IV. from the classical expression $H = \mathbf{p}^2/2m + V(r)$.
2. For a concise treatment of the special theory of relativity and the Lorentz transformations, see H. Goldstein, *Classical Mechanics* (Addison-Wesley, Massachusetts 1959), Chapter 6.
3. Schrödinger, E. *Ann. Physik*, **81**, 109 (1926).
4. Klein, O. Z. *Physik*, **37**, 895 (1926): Fock, V. Z. *Physik.* **38**, 242 (1926) and **39**, 226 (1926); Gordon, W. Z. *Physik*, **40**, 117 (1926) and **40**, 121 (1926).
5. In this Chapter as well as the next, we use bold face to denote 3-vectors while 4-vectors will be represented by the same letter in italics. Also, the components of a 4-vector will be distinguished by means of Greek subscripts whereas Latin subscripts will denote the components of a 3-vector. Latin subscripts are also used when we want to refer specifically to the space components of a 4-vector.

where E is the energy of the particle and c is the velocity of light in vacuum. This fact is expressed by writing $p = \{\mathbf{p}, (i/c)E\}$. Similarly, the space-time coordinate 4-vector $x = (\mathbf{r}, ict) = (\mathbf{x}, ict)$. Thus Eq. (10.1) leads to the identification,

$$\hat{E} = i\hbar \frac{\partial}{\partial t}, \tag{10.1a}$$

in addition to (3.18[1]) which may be written as

$$\cdot\hat{\mathbf{p}} = (\hbar/i)\nabla. \tag{10.1b}$$

(10.1a) could have been inferred from Eqs. (1.16) as suggested at the end of chapter 1, but the present derivation shows that the identifications (10.1a, b) are Lorentz covariant thus establishing their validity in a relativistic theory.

We should caution here that the generalisation (10.1) and its offshoot (10.1a) should be regarded as of a formal nature valid for use in a wave equation but of no deeper significance. For example, a similar generalisation of Eq. (3.18[2]) and the implied existence of a time operator are not intended (See, Section 3.2. *Time-energy uncertainty relationship*). It is to emphasize this formal aspect of Eq. (10.1a) that we have used the symbol $\hat{E}$ rather than $\hat{H}$. The latter will denote the Hamiltonian of the system in terms of the co-ordinates, momenta, etc.

In terms of $\hat{E}$ and $\hat{\mathbf{p}}$, the quantum mechanical wave equation is given by (cf. Eq. (4.14b)),

$$\hat{E}\psi(\mathbf{r}, t) = \hat{H}\psi(\mathbf{r}, t). \tag{10.3}$$

The Schrödinger equation for a free particle results from substituting for $\hat{H}$ the expression based on the non-relativistic formula (here, m is the mass of the particle),

$$H = \mathbf{p}^2/2m. \tag{10.4[1]}$$

The result is the equation,

$$i\hbar\partial_t \psi(\mathbf{r}, t) = -(\hbar^2/2m)\, \nabla^2\psi(\mathbf{r}, t), \tag{10.5}$$

where

$$\partial_t \equiv \frac{\partial}{\partial t}.$$

An obvious way to make Eq. (10.3) relativistic would be to use, in place of (10.4[1]), the relativistic expression for H, namely,

$$H = \sqrt{\mathbf{p}^2c^2 + m^2c^4}. \tag{10.4[2]}$$

However, this expression introduces the difficulty of defining the square root of a linear operator. In order to avoid this difficulty, $\hat{H}$ is squared before introducing into the wave equation. This requires $\hat{E}$ also to be squared so that the wave equation becomes [in place of (10.3)],

$$\hat{E}^2\psi(\mathbf{r}, t) = \hat{H}^2\psi(\mathbf{r}, t)$$

$$= c^2(\hat{\mathbf{p}}^2 + m^2c^2)\psi(\mathbf{r}, t), \tag{10.6[1]}$$

or,

$$(\Box - \kappa^2)\phi(x) = 0. \tag{10.6[2]}$$

Here,

$$\kappa = (mc/\hbar), \tag{10.7}$$

while $\Box$ is the *d'Alembertian given* by

$$\Box = \partial_\mu \partial_\mu = \nabla^2 - (1/c^2)\partial_t^2, \tag{10.8}$$

with

$$\partial_\mu \equiv \frac{\partial}{\partial x_\mu}.$$

Also, the Einstein summation convention is used (that is, a repeated (Greek) index is summed over from 1 to 4).

Eq. (10.6^2) is the Klein-Gordon equation. The symmetrical way in which space and time coordinates occur in this equation (it is second order in both, unlike Eq. (10.5) which is first order in time and second order in space) makes it manifestly Lorentz convariant. However, the act of squaring $\hat{H}$ has introduced two important new elements into the theory. One is that the wave equation no longer conforms to the dynamical postulate of chapter 4, which requires the equation to be first order in time. The other is the possibility of *negative energy* (through the negative square root of $\hat{H}^2$). These new elements give rise to difficulties in the interpretation of the Klein-Gordon equation as a quantum mechanical wave equation, at least within the conceptual framework of the Schrödinger equation. For example, the probabilistic (ensemble) interpretation of quantum mechanics is based on the circumstance that the Schrödinger equa- tion permits the definition of a *positive-definite* probability density ρ and a probability current density $\mathbf{j}$ satisfying the equation of continuity (see Problem 4.1),

$$\partial_t \rho + \text{div}\, \mathbf{j} = 0, \tag{10.9^1}$$

or, in covariant form,

$$\partial_\mu j_\mu(x) = 0, \tag{10.9^2}$$

where,

$$j = (\mathbf{j}, ic\rho). \tag{10.10}$$

The equation of continuity ensures that the total number of particles (in the ensemble) is conserved (Problem 4.2).

The Klein-Gordon equation yields the 4-current density,

$$j_\mu(x) = (\hbar/2mi)\,[\phi^*(x)\partial_\mu\phi(x) - \phi(x)\partial_\mu\phi^*(x)]. \tag{10.11}$$

The corresponding probability density is,

$$\rho(\mathbf{r}, t) = (i\hbar/2mc^2)(\phi^*\partial_t\phi - (\partial_t\phi^*)\phi). \tag{10.11a}$$

Since the value of $\partial_t\phi$ can be prescribed independently of ϕ [a consequence of the second order character of Eq. (10.6^2)] a negative value for ρ cannot be ruled out.

Problem 10.1: Verify Eq. (10.11).

Because of the above problem of the negative probability density, the Klein-Gordon equation failed to win recognition as a correct relativistic generalisation

of the Schrödinger equation. Meanwhile, Dirac[6] succeeded in obtaining a first order relativistic wave equation (remember that the troubles of the Klein-Gordon equation were attributed to the squaring of $\hat{H}$ and $\hat{E}$) which not only yielded a positive-definite probability density but from which also emerged properties like the spin and the magnetic moment of the electron and the fine structure of the hydrogen atom. The negative energies, however, persisted, but even these were transformed by Dirac into one of the greatest triumphs of the theory by a re-interpretation[7] (the hole theory) which introduced for the first time in physics the concepts of antiparticles, pair creation and annihilation and the vacuum (of elementary particles).

But along with its successes, the Dirac theory also brought on the realisation that the conceptual frame-work of the nonrelativistic quantum mechanics is too restrictive for the development of a self-consistent relativistic quantum theory. For example, at relativistic velocities the kinetic energy of a particle may become comparable or even surpass its rest energy. According to the principle of the mass-energy equivalence ($E = mc^2$) of the theory of relativity, creation of new particles is a distinct possibility at these energies. Obviously then, conservation of particle number cannot be a fundamental feature of relativistic quantum mechanics. On the other hand, the electric charge is a strictly conserved quantity. The equation of continuity of a relativistic wave equation should, therefore, refer to the conservation of the electric charge rather than to the conservation of particle-number. We note that if we multiply Eq. (10.11a) by e (a unit of charge) and interpret $(e\rho)$ as the *charge density*, the problem of its negative value disappears, as a charge density can be negative as well as positive.

The foregoing considerations led to a revival of the Klein-Gordon equation by Pauli and Weisskopf[8] who interpreted it as the (classical) *field equation* of spin-zero particles, the functions $\psi(\mathbf{r}, t)$ acting as *coordinates*[9] of the field. Similarly, the Dirac equation is the field equation of the spin-half particles. The quantization of these fields leads to the correct relativistic quantum theory of the associated particles. This does not, however, mean that the quantum field theory of particles is devoid of problems. In fact, many problems, mainly of a mathematical nature, are encountered in this theory. Nevertheless, its successes have been impressive enough to instill confidence in the correctness of its basic approach.

It is a basic feature of the relativistic quantum theory that particles with different spins are described by different wave equations. Thus, spin $\frac{1}{2}$ particles

---

6. Dirac, P.A.M. *Proc. Roy. Soc. (London)* **A 117**, 610 (1928).
7. Dirac, P.A.M. *Proc. Roy. Soc. (London)* **A 126**, 360 (1930).
8. Pauli, W. and Weisskopf, V. *Helv. Phys. Acta*, **7**, 709 (1934).
9. Note, in this connection, that the equations of motion of classical mechanics are second order in time for the coordinates. The equations of motion of the electromagnetic field are also second order in time. A field differs from a mechanical system mainly in having an infinite number of degrees of freedom. Since ψ at each space-time point represents an independent degree of freedom, the total number of degrees of freedom represented by ψ viewed as a coordinate is infinite.

are described by first order wave equations (the Dirac and the Weyl equations) whereas spin-zero and spin-one particles are described by second order wave equations (the Klein-Gordon equations and the wave equations of the electromagnetic field). This appears to be a consequence of the intimate relationship between spins and statistics (Section 9.2): It is found that a given (relativistic) wave equation can be consistently quantized using only one type of statistics (either the Bose-Einstein *or* the Fermi-Dirac), which type being decided by the equation. This is in contrast with the nonrelativistic Schrödinger equation which, viewed as a field equation, can be quantized using either of the statistics.

Eventhough a true relativistic quantum theory is, thus, a quantum field theory, it is sufficient for many purposes, especially for the fermions at low velocities, to follow the historical path of a quantum mechanical approach. This chapter is devoted to a discussion of the relativistic quantum mechanics while we will present an elementary introduction to the quantum field theory in the next chapter.

10.2 THE FIRST ORDER WAVE EQUATIONS

From Eqs. (10.1) and (10.3), it follows that a first order wave equation requires a Hamiltonian that is linear in the momentum vector. In the case of zero-mass particles, the classical Hamiltonian,

$$H = c \mid \mathbf{p} \mid, \tag{10.12}$$

suggests a quantum mechanical Hamiltonian of the form,

$$\hat{H} = c(\hat{\alpha} \cdot \hat{\mathbf{p}}), \tag{10.13^1}$$

where $\hat{\alpha}$ is a vector operator, *independent of the space and time co-ordinates*, such that

$$(\hat{\alpha} \cdot \hat{\mathbf{p}}) = \sum_{k=1}^{3} \hat{\alpha}_k \hat{p}_k = \sqrt{\sum_{k=1}^{3} \hat{p}_k^2}, \tag{10.14a}$$

Condition (10.14a) can be translated into conditions on the components of $\hat{\alpha}$ by squaring (10.14a). We have,

$$(\hat{\alpha} \cdot \hat{\mathbf{p}})^2 = \sum_{k=1}^{3} \hat{p}_k^2, \tag{10.15}$$

which requires,

$$\hat{\alpha}_i \hat{\alpha}_j + \hat{\alpha}_j \hat{\alpha}_i \equiv \{\alpha_i, \alpha_j\} = 2\delta_{ij}, (i, j = 1, 2, 3). \tag{10.16}$$

However, in this process we have admitted the negative square root in (10.14a):

$$(\hat{\alpha} \cdot \hat{\mathbf{p}}) = -\mid \mathbf{p} \mid, \tag{10.14b}$$

since (10.16) requires only the condition (10.15). Corresponding to (10.14b) the Hamiltonian is,

$$\hat{H} = -c(\hat{\alpha} \cdot \hat{\mathbf{p}}). \tag{10.13^2}$$

Now, relationships (10.16) are the same as the ones [Eq. (5.35b)] satisfied by the components of the Pauli spin vector $\hat{\sigma}$. Therefore, we can identify $\hat{\alpha}$ in this case, with $\hat{\sigma}$. In view of (10.13$^{1\text{-}2}$), the wave equation (10.3) then reads:

$$\hat{E}\psi = \pm c(\hat{\sigma} \cdot \hat{\mathbf{p}})\psi, \qquad (\hat{\sigma} \cdot \hat{\mathbf{p}}) = \pm \mid \hat{\mathbf{p}} \mid \tag{10.17}$$

or, in co-ordinate representation,

$$i\hbar\partial_t\Phi = -i\hbar c(\hat{\sigma}\cdot\nabla)\Phi, \quad (\hat{\sigma}\cdot\hat{\mathbf{p}}) = +|\hat{\mathbf{p}}|, \tag{10.17a}$$

and

$$i\hbar\partial_t\psi = i\hbar c(\hat{\sigma}\cdot\nabla)\psi, \quad (\hat{\sigma}\cdot\hat{\mathbf{p}}) = -|\hat{\mathbf{p}}|. \tag{10.17b}$$

These wave equations are known as the *Weyl equations*[10] of the neutrino. We will discuss them in Section 10.2B.

In the case of particles with non-zero mass, H is given by Eq. (10.4[2]) which can be put into a form resembling (10.12) by defining

$$p_0 = mc, \qquad \overset{o}{p} = (p_0, \mathbf{p}), \tag{10.18}$$

so that,

$$\left|\overset{o}{p}\right| = \sqrt{\sum_{\mu=0}^{3} p_\mu^2}. \tag{10.19}$$

Then,
$$H = c\left|\overset{o}{p}\right|, \tag{10.20}$$

and[11]

$$\hat{H} = c(\overset{o}{\alpha}\overset{o}{p}) = c\sum_{\mu=0}^{3}\hat{\alpha}_\mu\hat{p}_\mu, \tag{10.21^1}$$

with
$$(\overset{o}{\alpha}\overset{o}{p}) = \hat{\mathbf{p}}^2 + m^2c^2 \tag{10.22}$$

and,
$$\{\hat{\alpha}_\mu, \hat{\alpha}_\nu\} = 2\hat{\delta}_{\mu\nu}; \tag{10.23}$$

where,
$$\overset{\hat{o}}{\alpha} = (\hat{\alpha}_0, \hat{\alpha}). \tag{10.24}$$

Following convention, we put $\hat{\beta} = \hat{\alpha}_0$. Then (10.23) becomes equivalent to Eq. (10.16) plus the following:

$$\{\hat{\beta}, \hat{\alpha}_k\} = \hat{0}, (k = 1, 2, 3;) \tag{10.23a}$$

$$\hat{\beta}^2 = \hat{1}. \tag{10.23b}$$

In terms of $\hat{\beta}$ and $\hat{\alpha}$, the Hamiltonian is,

$$\hat{H} = c(\hat{\alpha}\cdot\hat{\mathbf{p}}) + \hat{\beta}mc^2. \tag{10.21^2}$$

The corresponding wave equation is

$$i\hbar\partial_t\psi = -i\hbar c(\hat{\alpha}\cdot\nabla)\psi + \hat{\beta}mc^2\psi. \tag{10.25^1}$$

This is known as the *Dirac equation*[6], while (10.21[2]) gives the *Dirac Hamiltonian.*

The above procedure of getting a Hamiltonian linear in the momentum vector by introducing certain operators 'from out of the blue' as it were, might look artificial. However, the procedure leads to useful results only because the operators so introduced are related to some property of the particles. In the case

---

10. Weyl, H. Z. *Physik*, **56**, 330 (1929).

11. Unlike in the case of mass-zero particles, the choice $\hat{H} = -c(\overset{o}{\alpha}\overset{o}{p})$ here, does not lead to any new results different from the one contained in the choice (10.21[1]).

of the zero-mass particles, the operator involved is the Pauli spin vector. The implication is that the associated particles have spin $\frac{1}{2}$. We will see that the operators α and β in the Dirac Hamiltonian are also related to the spin of the particles and to the existence of the antiparticles (negative energy states).

The Dirac equation was, historically, the first successful relativistic wave equation. The lessons learned from its success have been of great help in the understanding and the interpretation of other relativistic wave equations. It is only proper then to start the detailed discussion of the relativistic wave equations with that of the Dirac equation.

10.2A. The Dirac Equation

Since the operators $\hat{\alpha}$ and $\hat{\beta}$ are independent of the space and time co-ordinates and since they satisfy the anticommutation relationships (10.16) and (10.23a), it is clear that they cannot be represented either by algebraic numbers or by differential operators. Representation by matrices is the only alternative. We will, hereafter, regard them as matrices and drop the operator symbol when so considered. These matrices are referred to as the *Dirac matrices*.

In place of α and β, the set $\{\gamma_\mu\}$ defined below is also commonly used.

$$\gamma_k = -i\beta\alpha_k, (k = 1, 2, 3); \gamma_4 = \beta. \tag{10.26}$$

From (10.16) and (10.23), we get:

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}. \tag{10.27}$$

In terms of γ_μ, the Dirac equation (10.25^1) takes the form,

$$(\gamma_\mu \partial_\mu + \kappa)\psi(x) = 0, \tag{10.25^2}$$

where, $$\kappa = mc/\hbar.$$

Eq. (10.25^2) is the *covariant form* of the Dirac equation.

The properties of the Dirac equation are determined by the properties of the Dirac matrices. Let us, therefore, consider these properties.

Properties of the Dirac Matrices

DM 1. Since the Dirac Hamiltonian should be hermitian, it follows from (10.21^2) and (10.26) that α_k and β, as well as γ_μ, are *hermitian.*

DM 2. The eigenvalues of the matrices are ± 1.

DM 3. From $\gamma_\mu\gamma_\nu = -\gamma_\nu\gamma_\mu$, we have, $\gamma_\mu = -\gamma_\nu\gamma_\mu\gamma_\nu$, so that

$$Tr(\gamma_\mu) = -Tr(\gamma_\nu\gamma_\mu\gamma_\nu) = -Tr(\gamma_\nu^2\gamma_\mu) = -Tr(\gamma_\mu),$$

where, we have made use of property (A.24a) of the trace of a matrix. It follows that

$$Tr(\gamma_\mu) = 0. \tag{10.28}$$

The matrices α_k and β also share this property.

DM 4. Combining properties DM2 and DM3, we see that the order of the γ_μ should be even, namely, 2, 4, 6,...

DM 5. Up to equivalence, the γ_μ's (and hence also α_k and β) have *only one* irreducible representation[12] and that is *of order* 4. The first part follows from a theorem of Frobenius in group theory, according to which the number n of inequivalent irreducible representations of a finite semi-simple algebra which possesses a unit element, is equal to the number of elements in the algebra that commute with all other elements in the algebra. The Dirac algebra consists of the 16 (independent) elements that can be formed out of the four γ_μ's. These are: $I(=\gamma_\mu^2)$, γ_μ, $-i\gamma_\mu\gamma_\nu(\mu \neq \nu)$, $i\gamma_\mu\gamma_\nu\gamma_\sigma(\mu \neq \nu \neq \sigma)$ and $\gamma_1\gamma_2\gamma_3\gamma_4 \equiv \gamma_5$. The only element that commutes with all other elements here is the unit element I.

The second part of property DM 5 follows from another result of group theory, namely,

$$N = d_1^2 + \ldots + d_n^2, \tag{10.29}$$

where N is the total number of elements in the algebra and d_r is the dimensionality of the rth irreducible representation.

DM 6. Since the matrices do not commute among themselves, only one of the matrices could be diagonal in any particular representation.

As an explicit case, we choose a representation in which $\gamma_4 = \beta$, is diagonal:

$$\beta = \gamma_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{10.30a}$$

$$\alpha_k = \begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix}, (k = 1, 2, 3) \tag{10.30b}$$

where σ_k are the Pauli spin matrices (5.34). From Eqs. (10.26) and (10.30a, b), we have,

$$\gamma_k = \begin{pmatrix} 0 & -i\sigma_k \\ i\sigma_k & 0 \end{pmatrix}, (k = 1, 2, 3). \tag{10.30c}$$

Since the Dirac matrices are of order 4, the Dirac equation will make sense only if the Dirac wavefunction is a column matrix with four rows. That is, $\psi(x)$ should have four components in the space of the γ_μ's.

---

12. A representation is *irreducible* if there exists no matrix that can transform, through a similarity transformation (see Section A5), *all* the representative matrices to block-diagonal form. Two representations are *equivalent* if they are related to each other by a similarity transformation.

$$\psi(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \\ \psi_3(x) \\ \psi_4(x) \end{pmatrix} \tag{10.31}$$

Substituting from Eqs. (10.30a, c) and (10.31) in the Dirac equation (10.25[2]), the latter reduces to a set of four coupled equations:

$$\begin{aligned} -(\partial_4 + \kappa)\psi_1 + i\partial_3\psi_3 + (i\partial_1 + \partial_2)\psi_4 &= 0, \\ -(\partial_4 + \kappa)\psi_2 + (i\partial_1 - \partial_2)\psi_3 - i\partial_3\psi_4 &= 0, \\ i\partial_3\psi_1 + (i\partial_1 + \partial_2)\psi_2 - (\partial_4 - \kappa)\psi_3 &= 0, \\ (i\partial_1 - \partial_2)\psi_1 - i\partial_3\psi_2 - (\partial_4 - \kappa)\psi_4 &= 0. \end{aligned} \tag{10.32}$$

Problem 10.2: Deduce (10.30c) as one of the possible choice that follow from (10.27) and (10.30a). What is the other choice?

Problem 10.3: If $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ are operators whose components commute with those of $\hat{\alpha}$, deduce

$$(\hat{\alpha}\cdot\hat{\mathbf{A}})(\hat{\alpha}\cdot\hat{\mathbf{B}}) = (\hat{\mathbf{A}}\cdot\hat{\mathbf{B}}) + i\hat{\sigma}^{D}.(\hat{\mathbf{A}}\times\hat{\mathbf{B}})$$

where σ^{D} is defined by (10.39).

The Free Dirac Particles

A four-component wavefunction signifies four degrees of freedom. In order to ascertain what these are, let us consider the case of the free particle whose Hamiltonian is given by Eq. (10.21[2]).

For a free particle, we expect the angular momentum to be a constant of motion. This means that the components $\hat{L}_k$ (k = 1, 2, 3) given by Eq. (5.2a) should commute with $\hat{H}$. Now,

$$\begin{aligned} [\hat{H}, \hat{L}_1] &= \left[c\sum_k \hat{\alpha}_k \hat{p}_k + \beta mc^2, \hat{x}_2\hat{p}_3 - \hat{x}_3\hat{p}_2\right] \\ &= c\sum_k \hat{\alpha}_k \{[\hat{p}_k, \hat{x}_2\hat{p}_3] - [\hat{p}_k, \hat{x}_3\hat{p}_2]\} \\ &= c\sum_k \hat{\alpha}_k \{[\hat{p}_k, \hat{x}_2]\hat{p}_3 - [\hat{p}_k, \hat{x}_3]\hat{p}_2\} \\ &= -i\hbar c(\alpha_2\hat{p}_3 - \alpha_3\hat{p}_2) = -i\hbar c(\hat{\alpha}\times\hat{\mathbf{p}})_1, \end{aligned} \tag{10.33a}$$

where the relationships, $[\hat{x}_i, \hat{p}_j] = i\hbar\delta_{ij}$ are used.

Hence, $$[\hat{H}, \hat{\mathbf{L}}] = i\hbar c(\hat{\mathbf{p}}\times\hat{\alpha}) \neq 0. \tag{10.33b}$$

Thus, the angular momentum associated with the orbital motion of the particle is *not* a constant of motion[13] for the Dirac particle. But, the operator corresponding to the total angular momentum of a system is the generator of infinitesimal rotations of the system [see the remarks following Eq. (5.134a)]. Therefore, invariance of the system under rotations (which is a must from the viewpoint of the theory of relativity) requires that the angular momentum operator should commute with the Hamiltonian. The only possible conclusion is that the total angular momentum of the particle is not represented by $\hat{\mathbf{L}}$ but by an operator $\hat{\mathbf{J}}$ given by

$$\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}, \tag{10.34}$$

where $\hat{\mathbf{S}}$ should satisfy the following conditions:

(1) The components of $\hat{\mathbf{S}}$ should satisfy the basic commutation rules (cf. (5.4)),

$$[\hat{S}_i, \hat{S}_j] = i\hbar \in_{ijk} \hat{S}_k, \tag{10.35a}$$

that are obeyed by the components of an angular momentum operator.

(2) Components of $\hat{\mathbf{S}}$ should commute with the components of $\hat{\mathbf{L}}$ so that the usual rules of angular momentum addition can be used to obtain $\hat{\mathbf{J}}$ from $\hat{\mathbf{L}}$ and $\hat{\mathbf{S}}$.

(3) $\hat{\mathbf{J}}$ should commute with $\hat{H}$. This requires,

$$[\hat{H}, \hat{\mathbf{S}}] = -i\hbar c(\hat{\mathbf{p}} \times \hat{\boldsymbol{\alpha}}). \tag{10.35b}$$

(4) $\hat{\mathbf{S}}$ should not depend on the state of motion (that is, on $\mathbf{p}$). In other words, $\hat{\mathbf{S}}$ should depend only on $\hat{\boldsymbol{\alpha}}$ (since that is the only other vector operator in the theory).

It is possible to derive from conditions (3) and (4) above the following expression for the components of $\hat{\mathbf{S}}$.

$$\hat{S}_i = -(i\hbar/4) \in_{ijk} \hat{\alpha}_j \hat{\alpha}_k = -(i\hbar/4) \in_{ijk} \hat{\gamma}_j \hat{\gamma}_k, (i = 1, 2, 3). \tag{10.36}$$

The $\in_{ijk}$ in (10.35a) and (10.36) is defined in Eq. (5.2b). Defining the *Dirac spin vector* σ^D by

$$\mathbf{S} = (\hbar/2)\sigma^D, \tag{10.37}$$

we have,

$$\hat{\sigma}_i^D = -(i/2) \in_{ijk} \hat{\alpha}_j \hat{\alpha}_k, \tag{10.38a}$$

$$\hat{\sigma}^D = -(i/2)(\hat{\alpha} \times \hat{\alpha}). \tag{10.38b}$$

In matrix representation, we get [using (10.30b) and (10.38b)],

$$\sigma^D = \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}. \tag{10.39}$$

where σ is the matrix (vector) whose components are the Pauli spin matrices.

From (10.36), we have,

$$\hat{S}_3^2 = -\frac{\hbar^2}{4}(\hat{\alpha}_2\hat{\alpha}_3)^2 = \frac{\hbar^2}{4}\hat{\alpha}_2^2\hat{\alpha}_3^2 = (\hbar^2/4)\hat{1}.$$

---

13. We will see that, in the nonrelativistic limit, $\hat{\alpha}$ reduces to the velocity operator (hence parallel to $\hat{\mathbf{p}}$) so that in that limit $(\hat{\mathbf{p}} \times \hat{\alpha}) = \hat{0}$ and $\mathbf{L}$ becomes a constant of motion.

Thus, the eigenvalues of $\hat{S}_3$ (as also of $\hat{S}_1$ and $\hat{S}_2$) are $\pm(\hbar/2)$. We also verify that conditions (10.35a) are satisfied. Therefore, $\hat{\mathbf{S}}$ could be interpreted as an angular momentum of the particles. Since, however, $\hat{\mathbf{S}}$ does not arise from the orbital motion, it is called the *intrinsic* angular momentum, or *spin*, of the particles.[14] We conclude that the Dirac particles have spin $\frac{1}{2}$.

Problem 10.4: Deduce (10.36) from Eq. (10.35b).

Problem 10.5: Verify the following properties of the components of $\hat{\sigma}^D$.

$$\hat{\sigma}_i^D \hat{\sigma}_j^D = i \in_{ijk} \hat{\sigma}_k^D \ (i \neq j)$$

$$\{\hat{\sigma}_i^D, \hat{\sigma}_j^D\} = 2\hat{\delta}_{ij} \ (i, j = 1, 2, 3).$$

Now, a spin $\frac{1}{2}$ particle has only two degrees of freedom, namely, spin-up and spin-down ($S_3 = +\hbar/2$ and $S_3 = -\hbar/2$). We have to account for the additional two degrees of freedom of the Dirac particles. In view of (10.22) which does not exclude the possibility

$$\hat{\alpha}\hat{p} = -\sqrt{\mathbf{p}^2 + m^2c^2},$$

we surmise that the additional degrees of freedom might be related to the existence of negative energy states. That indeed is the case is confirmed by the following analysis.

A free-particle (plane-wave) solution of the Dirac equation can be written as

$$\psi(x) = u(\mathbf{p}) \exp[(i/\hbar)px] = u(\mathbf{p}) \exp[(i/\hbar)(\mathbf{p}\cdot\mathbf{x} - Et)], \tag{10.40}$$

where $u(\mathbf{p})$ is a 4-component *spinor*[15]:

$$u(\mathbf{p}) = \begin{pmatrix} u_1(\mathbf{p}) \\ u_2(\mathbf{p}) \\ u_3(\mathbf{p}) \\ u_4(\mathbf{p}) \end{pmatrix} \tag{10.41}$$

We choose the z-axis along the direction of motion of the particles. Then,

$$\psi_r(x) = \psi_r(x_3, t) = u_r(p_3) \exp[(i/\hbar)(p_3x_3 - Et)], \ (r = 1, 2, 3, 4). \tag{10.40a}$$

Substituting from (10.40a) in (10.32), we get

$$-cp_3u_1 + (E + mc^2)u_3 = 0,$$

$$(E - mc^2)u_1 - cp_3u_3 = 0, \tag{10.32^i}$$

14. The hypothesis of spin preceded the Dirac theory (See footnote 3, Chapter 5).
15. A spinor is a vector in the spin space. It is distinguished from an ordinary vector by its peculiar transformation properties under rotations [see Eq. (5.145c) and (10.135^3)]. We have already seen that the space of the Dirac matrices is, in fact, a spin space.

$$(E - mc^2)u_2 + cp_3u_4 = 0,$$

$$cp_3u_2 + (E + mc^2)u_4 = 0.$$

For a non-trivial solution of this set of equations, the determinant of the coefficient matrix should be zero (see Section A 6). That is,

$$(cp_3)^2 - (E^2 - m^2c^4) = 0,$$

or

$$E = \pm \epsilon_{p3}, \tag{10.42}$$

with

$$\epsilon_p = +\sqrt{\mathbf{p}^2c^2 + m^2c^4}. \tag{10.42a}$$

We might be tempted to throw away the negative energy[16] on the ground that it is unphysical. That we cannot do this will be clear from a consideration of the wavefunctions:

Case 1: $E = \epsilon_p > 0$.

Substituting ϵ_p for E in (10.32[1]), we get two independent solutions for the spinor u. These are,

$$u^{(1)} = \begin{pmatrix} a \\ 0 \\ b \\ 0 \end{pmatrix}; \quad u^{(2)} = \begin{pmatrix} 0 \\ a \\ 0 \\ -b \end{pmatrix}, \tag{10.43a}$$

where,

$$a = \frac{1}{\sqrt{2}}\left(1 + \frac{mc^2}{\epsilon_p}\right)^{1/2},$$

$$b = \frac{1}{\sqrt{2}}\left(1 - \frac{mc^2}{\epsilon_p}\right)^{1/2}. \tag{10.44}$$

Case 2: $E = -\epsilon_p < 0$.

In this case, we have,

$$u^{(3)} = \begin{pmatrix} -b \\ 0 \\ a \\ 0 \end{pmatrix}; \quad u^{(4)} = \begin{pmatrix} 0 \\ b \\ 0 \\ a \end{pmatrix}. \tag{10.43b}$$

16. The distinction between the negative value in (10.42) and the negative value in (4.117) in the case of the hydrogen atom, should be noted. In the former case, it is the total energy (including the rest energy) of the particle that is negative whereas, in the latter case, the negative energy (which excludes rest energy) merely indicates that the potential energy dominates over the kinetic energy.

Using (10.39), we verify,

$$\sigma_3^D u^{(1)} = +u^{(1)};$$

$$\sigma_3^D u^{(2)} = -u^{(2)}. \tag{10.45a}$$

$$\sigma_3^D u^{(3)} = +u^{(3)};$$

$$\sigma_3^D u^{(4)} = -u^{(4)}. \tag{10.45b}$$

Thus, $u^{(1)}$ and $u^{(2)}$ are, respectively, the spin-up and the spin-down states with positive energy while $u^{(3)}$ and $u^{(4)}$ are similar states with negative energy.

Also,

$$\sum_{r=1}^{2} u^{(r)} u^{(r)\dagger} = \begin{pmatrix} a^2 & 0 & ab & 0 \\ 0 & a^2 & 0 & -ab \\ ba & 0 & b^2 & 0 \\ 0 & -ab & 0 & b^2 \end{pmatrix}, \tag{10.46a}$$

and

$$\sum_{r=1}^{4} u^{(r)} u^{(r)\dagger} = (a^2 + b^2) I = I, \tag{10.46b}$$

where, I is the unit matrix of order 4.

Comparing Eqs. (10.46a, b) with the condition (2.122b) for the completeness of a set of vectors, we see that the positive energy states alone do not constitute a closed Hilbert space whereas the positive and negative energy states together form such a closed space. As a result, even if we have initially a positive energy particle, it can make a transition to a negative energy state in the presence of an external field. It is clear, therefore, that the negative energy states are integral parts of the theory. This aspect also explains the four components of ψ as arising from the two spin states each corresponding to the two energy states.

In the non-relativistic limit, $\epsilon_p \approx mc^2$, so that, from (10.44), $a \approx 1, b \approx 0$. We then see from (10.43a) and (10.46a) that $u^{(1)}$ and $u^{(2)}$ could be regarded as essentially 2-component wavefunctions spanning a closed Hilbert space. Therefore, in this limit, a positive-energy particle will remain as a positive-energy particle even in the presence of interactions, and the problem of negative energies will not arise.

Problem 10.6: Verify Eqs. (10.43a, b),

The Equation of Continuity

Now that the Dirac equation is also found to be not free from the negative energies, we have to make sure that the equation, though first order in time, is not bedevilled (like the Klein-Gordon equation) with negative probability densities.

Taking the Hermitian conjugate of Eq. (10.25[2]), we have,

$$(\partial_\mu \psi)^\dagger \gamma_\mu + \kappa \psi^\dagger = 0.$$

Multiplying this equation from the right by γ_4, we get, since

$$\partial_k^* = \partial_k, \partial_4^* = -\partial_4,$$

$$\partial_\mu \bar{\psi}\gamma_\mu - \kappa\bar{\psi} = 0, \tag{10.47}$$

where, $$\bar{\psi} = \psi^\dagger\gamma_4, \tag{10.48}$$

is known as the *Dirac adjoint of* ψ, while (10.47) is the *adjoint Dirac equation*. Using the representation (10.30a) and (10.31), we have,

$$\bar{\psi} = \overbrace{\psi_1^*\psi_2^* - \psi_3^* - \psi_4^*}. \tag{10.48a}$$

Left-multiplying (10.25[2]) by $\bar{\psi}$, right-multiplying (10.47) by ψ and adding the two, we get,

$$\partial_\mu(\bar{\psi}\gamma_\mu\psi) = 0. \tag{10.49}$$

Now, $\bar{\psi}\gamma_\mu\psi = \psi^\dagger\gamma_4\gamma_\mu\psi$, has the dimension of a probability density. Therefore, $c\bar{\psi}\gamma_\mu\psi$ has the dimension of a probability current density. But $c\bar{\psi}\gamma_\mu\psi$ is antihermitian. The 4-vector,

$$j_\mu = ic\bar{\psi}\gamma_\mu\psi, \tag{10.50}$$

is then seen to have the right form to be a probability current density. Multiplying Eq. (10.49) by ic, we get the equation of continuity (10.9[2]) with the probability current j_μ given by Eq. (10.50). The probability density ρ is defined by

$$j_4 = ic\rho, \text{ or } \rho = -(i/c)j_4 = \psi^\dagger\psi. \tag{10.51}$$

Thus, the probability density associated with the Dirac equation is positive definite and so there is no difficulty in interpreting the Dirac wavefunction ψ as a probability amplitude.

Non-relativistic Limit

Next, we should verify that the Dirac equation has a sensible nonrelativistic limit. We have already seen from Eqs. (10.43a), (10.44) and (10.46a), that, in the case of free particles, a 2-component description (corresponding to the two spin states of a spin $\frac{1}{2}$ particle) appears to be sufficient in the non-relativistic limit. We have to show that such a description is adequate even when the particle is not free. For this, we consider a Dirac particle in an external electromagnetic field described by the vector potential $\mathbf{A}$ and the scalar potential Φ [see Eq. (8.184)]. If e denotes the electric charge of the Dirac particle, the effect of the field on the particle would be to change its momentum from $\mathbf{p}$ to $\mathbf{P} = \mathbf{p} - (e/c)\,\mathbf{A}$ and the energy from E to $E - e\Phi$. Correspondingly, the Dirac equation (10.25[1]) becomes (where, $E\psi = i\hbar\partial_t\psi$),

$$(E - e\Phi)\psi = c(\hat{\alpha}\cdot\hat{\mathbf{P}})\psi + \beta mc^2\psi. \tag{10.52}$$

Writing,

$$\psi = \begin{pmatrix} \phi \\ \chi \end{pmatrix}, \tag{10.53}$$

where,

$$\phi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}; \chi = \begin{pmatrix} \psi_3 \\ \psi_4 \end{pmatrix}, \tag{10.54}$$

and substituting for $\hat{\alpha}$ and $\hat{\beta}$ from (10.30a, b), Eq. (10.52) splits up into the following two equations:

$$(E - e\Phi)\phi = c(\hat{\sigma} \cdot \hat{\mathbf{P}})\chi + mc^2\phi, \tag{10.52a}$$

$$(E - e\Phi)\chi = c(\hat{\sigma} \cdot \hat{\mathbf{P}})\phi - mc^2\chi. \tag{10.52b}$$

Writing, $E = E' + mc^2$, we get from (10.52b),

$$\chi = \left\{ \frac{c(\hat{\sigma} \cdot \hat{\mathbf{P}})}{E' - e\Phi + 2mc^2} \right\} \phi, \tag{10.55a}$$

Similarly, putting $E = E' - mc^2$ in Eq. (10.52a) we get,

$$\phi = \left\{ \frac{c(\hat{\sigma} \cdot \hat{\mathbf{P}})}{E' - e\Phi - 2mc^2} \right\} \chi. \tag{10.55b}$$

In the non-relativistic case, $| E' - e\Phi | \ll 2mc^2$ and $| \mathbf{P} | \ll mc$, so that the following approximations could be made [Remember that $(E' - e\Phi)$ is the kinetic energy and $\mathbf{P} \approx m\mathbf{v}$, is kinetic momentum] :

$$\chi \approx \frac{(\hat{\sigma} \cdot \mathbf{P})}{2mc} \phi \ll \phi \quad \text{(positive energy)} \tag{10.56a}$$

and

$$\phi \approx -\frac{(\hat{\sigma} \cdot \mathbf{P})}{2mc} \chi \ll \chi \text{ (negative energy).} \tag{10.56b}$$

Thus, in the case of positive energy states, ϕ represents the *large components* while χ denotes the *small components.*

Substituting from (10.56a) in (10.52a), we have,

$$(E - mc^2)\phi \approx \left[\frac{(\hat{\sigma} \cdot \hat{\mathbf{P}})^2}{2m} + e\Phi \right] \phi. \tag{10.57[1]}$$

Now,

$$(\hat{\sigma} \cdot \hat{\mathbf{P}})^2 = (\hat{\sigma} \cdot \hat{\mathbf{P}})(\hat{\sigma} \cdot \hat{\mathbf{P}}) = \hat{\mathbf{P}}^2 + i\hat{\sigma}.(\hat{\mathbf{P}} \times \hat{\mathbf{P}}), \text{ (by problem 5.6)}$$

$$= \left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A} \right)^2 - (e\hbar/c)(\hat{\sigma} \cdot \hat{\mathbf{B}}), \tag{10.58}$$

Since

$$(\hat{\mathbf{P}} \times \hat{\mathbf{P}}) = \left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A} \right) \times \left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A} \right)$$

$$= -(e/c)[\mathbf{A} \times \hat{\mathbf{p}} + \hat{\mathbf{p}} \times \mathbf{A}] = (-e/c)(-i\hbar\mathbf{B}) \tag{10.59}$$

where, $\mathbf{B} = \nabla \times \mathbf{A}$, is the magnetic field (see Eq. (8.184)).
Here, the commutation rules (3.14d) are used.

In the non-relativistic case,

$$(E - mc^2)\phi = i\hbar\partial_t\phi. \tag{10.60}$$

Thus, (10.57[1]) reduces to :

$$i\hbar\partial_t\phi = \left\{\frac{(\hat{\mathbf{p}} - (e/c)\mathbf{A})^2}{2m} - \frac{e\hbar}{2mc}(\hat{\boldsymbol{\sigma}}\cdot\mathbf{B}) + e\Phi\right\}\phi. \tag{10.57^2}$$

This is the *Pauli equation*[17] for the electron. The second term on the R.H.S. is the potential energy associated with a magnetic dipole of moment,

$$\mathbf{M} = \left(\frac{e\hbar}{2mc}\right)\boldsymbol{\sigma}. \tag{10.61}$$

Since the eigenvalues of $\hat{\sigma}$ along any direction are ± 1, we have,

$$\mathbf{M}\ (\text{observed}) = \langle \hat{M}_z \rangle = \pm\mu_B, \tag{10.61a}$$

where $\mu_B = (e\hbar/2mc)$, is known as the *Bohr magneton*.

Also, since e is negative for the electron, $\mathbf{M}$ is antiparallel to the spin. These results are in agreement with the observed magnetic moment of the electron. Thus, Eq. (10.57[2]) not only establishes a meaningful non-relativistic limit for the Dirac equation, but also shows that the Dirac particles (at least the positive energy ones) are electrons.

Problem 10.7: Show that in the case of a weak magnetic field, Eq. (10.57[2]) can also be written as

$$i\hbar\partial_t\phi = \left[\frac{\mathbf{p}^2}{2m} - \frac{e}{2mc}(\hat{\mathbf{L}} + 2\hat{\mathbf{S}})\cdot\mathbf{B}\right]\phi,$$

where, $\hat{\mathbf{L}}$ is the orbital angular momentum operator.

According to (10.56) and (10.53), $\phi(x)$ and $\chi(x)$ could be regarded as the non-relativistic limits of $\psi(x)$ for the positive and the negative energy states, respectively. Here, ϕ and χ are 2-component wavefunctions whereas ψ is a 4-component one. We can make use of these functions to define the *non-relativistic equivalent* (or limit) $\hat{\Omega}_{nr}$ of a Dirac operator $\hat{\Omega}\cdot\hat{\Omega}_{nr}$ is the non-relativistic equivalent of $\hat{\Omega}$ if $\langle\psi_b \mid \hat{\Omega} \mid \psi_a\rangle$ approaches $\langle\phi_b \mid \hat{\Omega}_{nr} \mid \phi_a\rangle$ for positive energy and $\langle\chi_b \mid \hat{\Omega}_{nr} \mid \chi_a\rangle$ for negative energy, in the non-relativistic limit.

Define,

$$\hat{\pi}_\pm = \frac{1}{2}(\hat{1} \pm \hat{\beta}). \tag{10.62}$$

Then,

$$\hat{\pi}_\pm^2 = \hat{\pi}_\pm;\ \hat{\pi}_+\hat{\pi}_- = \hat{0}. \tag{10.63}$$

---

17. Pauli, W. *Handbuch der Physik*, 2nd ed. Vol. 24, p. 1 (Springer, Berlin 1933). Note that ϕ is normalized only up to zero-order in the velocity [see Eq. (10.84)].

Thus, $\hat{\pi}_+$ and $\hat{\pi}_-$ are projection operators that projects on to the subspace corresponding to the eigenvalues +1 and −1, respectively, of $\hat{\beta}$. That is,

$$\hat{\pi}_+\psi = \psi_+; \hat{\pi}_-\psi = \psi_-. \tag{10.64}$$

where,

$$\beta\psi_\pm = \pm\psi_\pm,$$

or,

$$\beta\hat{\pi}_\pm = \pm\hat{\pi}_\pm. \tag{10.65}$$

Obviously,

$$\psi = \psi_+ + \psi_-, \ .$$

since

$$\hat{\pi}_+ + \hat{\pi}_- = \hat{1}. \tag{10.66}$$

Using representation (10.30a) and definition (10.54), we see,

$$\psi_+ = \begin{pmatrix} \phi \\ 0 \end{pmatrix}; \ \psi_- = \begin{pmatrix} 0 \\ \chi \end{pmatrix}. \tag{10.67}$$

In the limit of small velocities, we have,

$$\psi \underset{v\to 0}{\sim} \psi_+ \ (E > 0); \psi \underset{v\to 0}{\sim} \psi_- \ (E - 0). \tag{10.68}$$

Problem 10.8: Using (10.56a, b) show that, in the nonrelativistic limit,

$$\psi_- = [(\hat{\alpha}\cdot v)/2c]\psi_+ \text{ (positive energy),}$$

and

$$\psi_+ = -[(\hat{\alpha}\cdot v)/2c]\psi_- \text{ (negative energy).}$$

It is convenient to divide the operators of the Dirac theory into two classes depending on whether they commute or anticommute with $\hat{\beta}$. The former type are called *even* operators while the latter are called *odd* operators.

$$\beta\hat{\Omega}_e = \hat{\Omega}_e\beta; \beta\Omega_0 = -\Omega_0\beta. \tag{10.69}$$

Also,

$$\hat{\Omega}_e\hat{\pi}_\pm = \hat{\pi}_\pm\hat{\Omega}_e; \hat{\Omega}_0\hat{\pi}_\pm = \hat{\pi}_\mp\hat{\Omega}_0. \tag{10.70}$$

An arbitrary operator $\hat{\Omega}$ can be written as sum of an even part and an odd part:

$$\hat{\Omega} = \hat{\Omega}_e + \hat{\Omega}_0, \tag{10.71}$$

where,

$$\hat{\Omega}_e = \frac{1}{2}(\hat{\Omega} + \beta\hat{\Omega}\beta); \hat{\Omega}_0 = \frac{1}{2}(\hat{\Omega} - \beta\hat{\Omega}\beta). \tag{10.71a}$$

We need, therefore, consider the non-relativistic equivalents of even and odd operators only.

$$\langle \psi_b \mid \hat{\Omega}_e \mid \psi_a \rangle = \langle \psi_b \mid (\hat{\pi}_+ + \hat{\pi}_-)\hat{\Omega}_e(\hat{\pi}_+ + \hat{\pi}_-) \mid \psi_a \rangle$$

$$= \langle \psi_{b+} \mid \hat{\Omega}_e \mid \psi_{a+} \rangle + \langle \psi_{b-} \mid \hat{\Omega}_e \mid \psi_{a-} \rangle, \tag{10.72}$$

where Eqs. (10.70), (10.63) and (10.64) have been used.

Similarly,

$$\langle \psi_b \mid \hat{\Omega}_0 \mid \psi_a \rangle = \langle \psi_{b-} \mid \hat{\Omega}_0 \mid \psi_{a+} \rangle + \langle \psi_{b+} \mid \hat{\Omega}_0 \mid \psi_{a-} \rangle \tag{10.73}$$

Thus, an even operator connects components of like size while an odd operator connects the large components with the small components.

Now, (10.69) requires $\hat{\Omega}_e$ and $\hat{\Omega}_0$ to be of the following form:

$$\hat{\Omega}_e = \begin{pmatrix} \hat{\Omega}_e^{(+)} & 0 \\ 0 & \hat{\Omega}_e^{(-)} \end{pmatrix}; \hat{\Omega}_0 = \begin{pmatrix} 0 & \hat{\Omega}_0^{(+)} \\ \hat{\Omega}_0^{(-)} & 0 \end{pmatrix} \tag{10.74}$$

where, $\hat{\Omega}^{(\pm)}$ are 2×2 matrices. Then, using (10.67), we get

$$\langle \psi_{b+} \mid \hat{\Omega}_e \mid \psi_{a+} \rangle = \langle \phi_b \mid \hat{\Omega}_e^{(+)} \mid \phi_a \rangle, \tag{10.75a}$$

$$\langle \psi_{b-} \mid \hat{\Omega}_e \mid \psi_{a-} \rangle = \langle \chi_b \mid \hat{\Omega}_e^{(-)} \mid \chi_a \rangle, \tag{10.75b}$$

so that, from (10.72), we have,

$$\langle \psi_b \mid \hat{\Omega}_e \mid \psi_e \rangle \underset{v \ll c}{\approx} \langle \phi_b \mid \hat{\Omega}_e^{(+)} \mid \phi_a \rangle, (E > 0), \tag{10.76a}$$

and

$$\langle \psi_b \mid \hat{\Omega}_e \mid \psi_e \rangle \underset{v \ll c}{\approx} \langle \chi_b \mid \hat{\Omega}_e^{(-)} \mid \chi_a \rangle, (E < 0), \tag{10.76b}$$

Using the results of problem 10.8, we can write (10.73) as,

$$\langle \psi_b \mid \hat{\Omega}_0 \mid \psi_a \rangle \underset{v \ll c}{\approx} \pm \langle \psi_{b_\pm} \left| \left\{ \hat{\Omega}_0 \frac{(\hat{\alpha} \cdot \hat{v})}{2c} + \frac{(\hat{\alpha} \cdot \hat{v})}{2c} \hat{\Omega}_0 \right\} \right| \psi_{a_\mp} \rangle, \tag{10.77}$$

where, as usual, the upper sign corresponds to positive energy and the lower sign to negative energy. Again from (10.74), (10.67) and (10.30b) we get,

$$\langle \psi_{b+} \mid [\hat{\Omega}_0(\hat{\alpha} \cdot \hat{v})/2c] \mid \psi_{a+} \rangle = \langle \phi_b \mid [\hat{\Omega}_0^{(+)}(\hat{\sigma} \cdot \hat{v})/2c] \phi_a \rangle \tag{10.77a}$$

$$\langle \psi_b \mid [\hat{\Omega}_0(\hat{\alpha} \cdot \hat{v})/2c] \mid \psi_{a-} \rangle = \langle \chi_b \mid [\hat{\Omega}_0^{(-)}(\hat{\sigma} \cdot \hat{v})/2c] \chi_a \rangle \tag{10.77b}$$

so that

$$\langle \psi_b \mid \hat{\Omega}_0 \mid \psi_a \rangle \underset{v \ll c}{\approx} (1/2c) \langle \phi_b \mid \{ \hat{\Omega}_0^{(+)}(\hat{\sigma} \cdot \hat{v}) + (\hat{\sigma} \cdot \hat{v}) \hat{\Omega}_0^{(+)} \} \mid \phi_a \rangle,$$

$$\text{for } E > 0, \tag{10.78a}$$

and

$$\langle \psi_b \mid \hat{\Omega}_0 \mid \psi_a > \underset{v \ll c}{\approx} -(1/2c) \langle \chi_b \mid \{ \hat{\Omega}_0^{(-)}(\hat{\sigma} \cdot \hat{v}) + (\hat{\sigma} \cdot \hat{v}) \hat{\Omega}_0^{(-)} \} \mid \chi_a \rangle,$$

$$\text{for } E < 0. \tag{10.78b}$$

Thus, from (10.75) and (10.78), we have,

$$(\hat{\Omega}_e)_{nr} = \hat{\Omega}_e^{(+)}; \tag{10.79}$$

$$(\hat{\Omega}_0)_{nr} = \pm (1/2c) [\hat{\Omega}_0^{(+)}(\hat{\sigma} \cdot \hat{v}) + (\hat{\sigma} \cdot \hat{v}) \hat{\Omega}_0^{(+)}]. \tag{10.80}$$

Let us consider a few illustrative examples. As even operators, we choose $\hat{\beta}$ and $\hat{\sigma}^D$. From (10.30a), (10.39) and (10.79), we have,

$$\hat{\beta}_{nr} = \pm\hat{1}, \tag{10.79^1}$$

$$(\hat{\sigma}^D)_{nr} = \hat{\sigma}. \tag{10.79^2}$$

$\hat{\alpha}$ is an example of an odd operator. From (10.30b) and (10.80), we get,

$$(\hat{\alpha})_{nr} = \pm(1/2c)\,[\hat{\sigma}(\hat{\sigma}\cdot\hat{\mathbf{v}}) + (\hat{\sigma}\cdot\hat{\mathbf{v}})\hat{\sigma}]$$

$$= \pm(\mathbf{v}/c). \tag{10.80^1}$$

That is, at non-relativistic velocities, $c\hat{\alpha}$ reduces to the velocity operator. We see then from (10.33b) that, in this case, the orbital angular momentum becomes a constant of motion. This is not due to the disappearance of the spin (the spin is very much there, as indicated by (10.79^2) as well as by (10.57^2)), but due to the disappearance of the spin-orbit coupling which, unlike the spin, is truly of relativistic origin (see the next sub-section).

Spin-Orbit Coupling

In order to exhibit the presence of the spin-orbit interaction and to establish its relativistic origin, we consider an approximation which is a step higher than that which led to the Pauli equation (10.57^2). From (10.55a) and (10.56a), we see that the latter is valid up to zero-order[18] in (v^2/c^2). Therefore, we want the present approximation to be valid up to first order in (v^2/c^2). From (10.55a) we see that the required approximation, for the positive energy case, is:

$$\chi \approx \left(1 - \frac{E'-V}{2mc^2}\right)\frac{(\hat{\sigma}\cdot\hat{\mathbf{P}})}{2mc}\phi, \tag{10.81}$$

where $V(\mathbf{r}) = e\Phi(\mathbf{r})$.

Substituting (10.81) in (10.52a), we get (since $E = E' + mc^2$),

$$E'\phi = \hat{H}\phi, \tag{10.82^1}$$

with

$$\hat{H} = \frac{(\hat{\sigma}\cdot\hat{\mathbf{P}})}{2m}\left[1 - \frac{E'-V}{2mc^2}\right](\hat{\sigma}\cdot\hat{\mathbf{P}}) + \hat{V}(\mathbf{r})$$

$$= \left(1 - \frac{E'-V}{2mc^2}\right)\frac{\hat{\mathbf{P}}^2}{2m} + \hat{V}(\mathbf{r}) - \frac{e\hbar}{2mc}(\hat{\sigma}\cdot\mathbf{B}) + \frac{\hbar}{4m^2c^2}\{\hat{\sigma}\cdot(\nabla V\times\hat{\mathbf{p}})\} - \frac{i\hbar}{4m^2c^2}(\nabla V\cdot\hat{\mathbf{p}}), \tag{10.83^1}$$

---

18. $E' - e\Phi = mc^2\left[(1 - v^2/c^2)^{-1/2} - 1\right] = mc^2\left[\frac{1}{2}(v^2/c^2) + (3/8)(v^2/c^2)^2 + \ldots\right]$

Thus,

$$(E' - e\Phi)/mc^2 = 0\cdot(v^2/c^2)^0 + \frac{1}{2}\cdot(v^2/c^2)^1 + (3/8)(v^2/c^2)^2 + \ldots$$

to first-order in (v^2/c^2). Here, Eq. (10.58) as well as the relationship,

$$(\hat{\sigma}\cdot\hat{\mathbf{P}})\hat{F}(\mathbf{r})(\hat{\sigma}\cdot\hat{\mathbf{P}}) = \hat{F}(\mathbf{r})(\hat{\sigma}\cdot\hat{\mathbf{P}})^2 - i\hbar\cdot(\hat{\sigma}\cdot\nabla\hat{F})(\hat{\sigma}\cdot\hat{\mathbf{p}}),$$

which follows from an application of Eq. (3.14d), have been used. Now, in the case of the Pauli equation, the normalization of ϕ is given by

$$\begin{aligned} 1 &= \int \psi^+\psi d^3\mathbf{r} = \int(\phi^+\phi + \chi^+\chi)d^3\mathbf{r} \\ &\approx \int\left\{1 + \frac{(\hat{\sigma}\cdot\hat{\mathbf{P}})^2}{4m^2c^2}\right\}\phi^+\phi d^3\mathbf{r}, \qquad \text{(by (10.56a))} \\ &\approx \int \phi^+\phi d^3, \quad \text{(to zero-order in } (v^2/c^2)\text{).} \end{aligned}$$

That is, ϕ is normalized to the same approximation as the equation.

In the present case, we have,

$$\begin{aligned} \int \psi^+\psi d^3\mathbf{r} &\approx \int\left(1 + \frac{(\hat{\sigma}\cdot\hat{\mathbf{P}})^2}{4m^2c^2}\right)\phi^+\phi d^3\mathbf{r} \\ &\approx \int\left(1 + \frac{\hat{\mathbf{p}}^2}{4m^2c^2}\right)\phi^+\phi d^3\mathbf{r}, \end{aligned}$$

to first-order in (v^2/c^2). Therefore, ϕ is not a normalized eigenvector, but the normalized (to order v^2/c^2) function ψ is given by

$$\psi = \left(1 + \frac{\hat{\mathbf{p}}^2}{4m^2c^2}\right)^{1/2}\phi \approx \hat{C}\phi, \tag{10.84}$$

where,

$$\hat{C} = \left(\hat{1} + \frac{\mathbf{p}^2}{8m^2c^2}\right). \tag{10.84a}$$

In terms of ψ, Eq. (10.82^1) can be written as

$$E'\hat{C}^{-1}\psi = \hat{H}\hat{C}^{-1}\psi,$$

or (multiplying both sides by $\hat{C}$),

$$E'\psi = \hat{H}'\psi, \tag{10.82^2}$$

where,

$$\hat{H}' = \hat{C}\hat{H}\hat{C}^{-1} = \left(\hat{1} + \frac{\mathbf{p}^2}{8m^2c^2}\right)\hat{H}\left(\hat{1} - \frac{\mathbf{p}^2}{8m^2c^2}\right)$$

From (10.83^1), we see that $\hat{C}$ need be considered different from unity only in the case of $\hat{V}(\mathbf{r})$. And in the case of $\hat{V}(\mathbf{r})$, we have,

$$\left(\hat{1}+\frac{\mathbf{p}^2}{8m^2c^2}\right)\hat{V}(\mathbf{r})\left(\hat{1}-\frac{\mathbf{p}^2}{8m^2c^2}\right)\approx\hat{V}(\mathbf{r})+\frac{1}{8m^2c^2}[\hat{\mathbf{p}}^2\hat{V}(\mathbf{r})-\hat{V}(\mathbf{r})\hat{\mathbf{p}}^2]$$

$$\approx\hat{V}(\mathbf{r})+\frac{1}{8m^2c^2}[-\hbar^2\nabla^2\hat{V}-2i\hbar(\nabla\hat{V}\cdot\hat{\mathbf{p}})].$$

Thus, up to first order in (v^2/c^2),

$$\hat{H}'=\frac{\left(\hat{\mathbf{p}}-\frac{e}{c}\mathbf{A}\right)^2}{2m}+e\Phi-\frac{e\hbar}{2mc}(\hat{\boldsymbol{\sigma}}\cdot\mathbf{B})-\frac{\hat{\mathbf{p}}^4}{8m^3c^2}+\frac{\hbar}{4m^2c^2}[\hat{\boldsymbol{\sigma}}\cdot(\nabla\hat{V}\times\hat{\mathbf{p}})]$$

$$-\frac{i\hbar}{2m^2c^2}(\nabla\hat{V}\cdot\hat{\mathbf{p}})-\frac{\hbar^2}{8m^2c^2}\nabla^2\hat{V}. \qquad (10.83^2)$$

The first three terms in $\hat{H}'$ are the same as the ones appearing in the non-relativistic equation (10.57^2). The remaining terms are relativistic corrections of order (v^2/c^2) [Remember that $|-i\hbar\nabla\hat{V}|=|\hat{\mathbf{p}}\hat{V}|$]. The term in $\hat{\mathbf{p}}^4$ results from writing.

$$\left(\frac{E'-\hat{V}}{2mc^2}\right)\cdot\frac{\hat{\mathbf{P}}^2}{2m}\approx\frac{1}{2mc^2}\left(\frac{\hat{\mathbf{P}}^2}{2m}\right)^2\approx\frac{1}{2mc^2}\left(\frac{\hat{\mathbf{p}}^2}{2m}\right)^2,$$

and is the lowest-order relativistic correction to the kinetic energy operator as seen by expanding the energy operator $\sqrt{\hat{\mathbf{p}}^2c^2+m^2c^4}$ in powers of $(\hat{\mathbf{p}}^2/m^2c^2)$. The fifth term in (10.83^2) represents the spin-orbit coupling. In the case of a spherically symmetric potential $(\hat{V}(\hat{\mathbf{r}})=\hat{V}(r))$, we have

$$\nabla\hat{V}=(\hat{\mathbf{r}}/r)\frac{\partial\hat{V}}{\partial r},$$

and then,

$$\hat{H}'_{\text{spin-orbit}}=\frac{\hbar}{4m^2c^2}\cdot\frac{1}{r}\frac{\partial\hat{V}}{\partial r}(\hat{\boldsymbol{\sigma}}\cdot\hat{\mathbf{L}}), \qquad (10.85)$$

where $\hat{\mathbf{L}}=\hat{\mathbf{r}}\times\hat{\mathbf{p}}$. The last two terms in $\hat{H}'$ seem to be related to the existence of negative energy states (in fact, to the phenomenon of *Zitterbewegung*—the rapid fluctuations in the coordinate of the electron over a distance of the order of the Compton wave length $(\hbar/mc)$. These fluctuations arise due to the mixing of positive and negative energy states[19]). The last term is usually referred to as the *Darwin term.*[20]

The Foldy-Wouthuysen Transformation

We have seen that, in the non-relativistic limit, a separation between the positive and the negative energy states can be achieved in the form of 2-component equations. We should expect such a separation possible, in the case of free Dirac

19. See Ref. 1, Sections 3.3 and 4.3.
20. Darwin, C.G. [*Proc. Roy. Soc.* A **118**, 634 (1928)] was the first to obtain it.

particles, even in the relativistic case. For, in the absence of interactions, a particle in a positive energy state will remain in the positive energy state. But a spin-$\frac{1}{2}$ particle with positive energy has only two degrees of freedom. There is a certain redundancy in the description of such particles in terms of 4-component wavefunctions. Therefore, an equivalent description in terms of 2-component wavefunctions must exist.

Let us assume, then, that there exists a unitary transformation $\hat{U}$ such that the transformed Hamiltonian $\hat{H}' = \hat{U}\hat{H}\hat{U}^{\dagger}$, where $\hat{H} = c(\hat{\alpha}\cdot\hat{\mathbf{p}}) + \beta mc^2$, satisfies the eigenvalue equation,

$$\hat{H}'\psi'_{\pm} = \pm|E|\psi'_{\pm} = \pm\epsilon_{\mathbf{p}}\psi'_{\pm}, \tag{10.86}$$

with $|E| = c\sqrt{\mathbf{p}^2 + m^2c^2}$. Here, $\psi'_{\pm}$ are essentially 2-component wavefunctions. In fact, they could be taken to be of the same form as the $\psi_{\pm}$ in (10.67). We see, then, from (10.65) that Eq. (10.86) is satisfied if

$$\hat{U}\hat{H}\hat{U}^{\dagger} = \hat{H}' = \beta\epsilon_{\mathbf{p}}. \tag{10.87^1}$$

That is, the transformation should eliminate the odd operator $\hat{\alpha}$ [which causes mixing between ψ_+ and ψ_- (see (Eq. (10.52))] from $\hat{H}$. The problem of finding such a transformation is analogous to that of finding a transformation that eliminates $\hat{\sigma}_x$ from a two-component spin Hamiltonian of the form,

$$\hat{\mathcal{H}} = b_x\hat{\sigma}_x + b_z\hat{\sigma}_z. \tag{10.88^1}$$

The transformation in the case of (10.88^1) is a rotation (of the co-ordinate system) about the y-axis in spin space through an angle θ_y given by $\tan\theta_y = b_x/b_z$ (See Fig. 10.1(a)). The unitary operator corresponding to this rotation is, according to Eq. (5.135),

$$\hat{U}_y(\theta_y) = \exp\{(i/2)\sigma_y\theta_y\} = \exp\left\{\left(\frac{1}{2}\right)\hat{\sigma}_z\hat{\sigma}_x\theta_y\right\}. \tag{10.89}$$

Referring to Fig. 10.1(a), we see that $\hat{\mathcal{H}}$ is transformed into

$$\hat{U}_y\hat{\mathcal{H}}\hat{U}_y^{\dagger} = \hat{\mathcal{H}}' = b'_z\hat{\sigma}_z = \sqrt{b_z^2 + b_x^2}\,\hat{\sigma}_z. \tag{10.88^2}$$

In the case of the Dirac Hamiltonian, we have $\hat{\sigma}, \beta, c\hat{\mathbf{p}}$ and mc^2 respectively in place of $\hat{\sigma}_x, \hat{\sigma}_z, b_x$ and b_z. The unitary operator corresponding to (10.89) is, therefore, given by

$$\hat{U} = e^{i\hat{S}}, \tag{10.90}$$

with

$$\hat{S} = \frac{-i}{2mc}\beta(\hat{\alpha}\cdot\hat{\mathbf{p}})\Theta, \tag{10.90a}$$

where, Θ is a function of $(|\mathbf{p}|/mc.)$ We show below that the transformation (10.90) indeed transforms $\hat{H}$ to the form (10.87) with a proper choice of Θ.

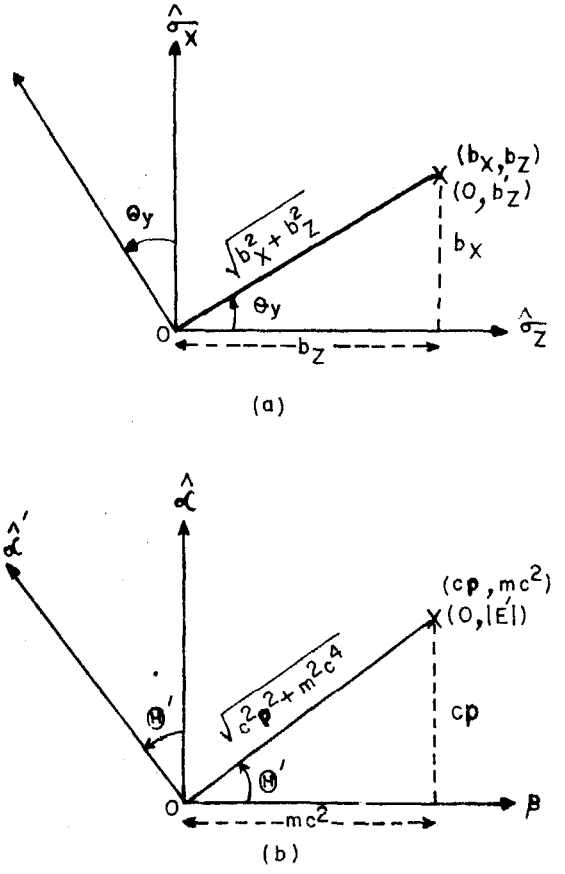


Fig. 10.1. Graphical representation of the Foldy-Wouthuysen transformation [In (b), $\Theta' = (|\mathbf{p}|/mc)\Theta$].

$$
\begin{aligned}
e^{iS} &= \exp\{(1/2mc)\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})\Theta\} \\
&= \hat{1} + \hat{\beta}\frac{(\hat{\alpha}\cdot\hat{\mathbf{p}})}{2mc}\Theta + \frac{1}{2!}\frac{\{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})\}^2}{(2mc)^2}\Theta^2 + \dots \\
&= \left[\hat{1} - \frac{1}{2!}\frac{|\mathbf{p}|^2\Theta^2}{(2mc)^2} + \frac{1}{4!}\frac{|\mathbf{p}|^4\Theta^4}{(2mc)^4} - \dots\right] \\
&+ \frac{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})}{|\mathbf{p}|}\left[\frac{|\hat{\mathbf{p}}|\Theta}{2mc} - \frac{1}{3!}\frac{(|\hat{\mathbf{p}}|\Theta)^3}{(2mc)^3} + \dots\right] \\
&= \cos\left(\frac{|\hat{\mathbf{p}}|\Theta}{2mc}\right) + \frac{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})}{|\mathbf{p}|}\sin\left(\frac{|\hat{\mathbf{p}}|\Theta}{2mc}\right),
\end{aligned}
\tag{10.91}
$$

since $$\{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})\}^2 = -\hat{\beta}^2\sum_k \hat{\alpha}_k^2\hat{p}_k^2 = -|\mathbf{p}|^2.$$

Also, $$(\hat{\alpha}\cdot\hat{\mathbf{p}})\{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})\} = -(\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})\}(\hat{\alpha}\cdot\hat{\mathbf{p}}), \tag{10.92a}$$

$$\hat{\beta}\{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})\} = -\{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})\}\hat{\beta}. \tag{10.92b}$$

Thus, $$\hat{H}\hat{S} = -\hat{S}\hat{H}, \tag{10.92c}$$

and $$\hat{H}e^{-i\hat{S}} = e^{i\hat{S}}\hat{H}.$$

Hence, $$\hat{H}' = e^{i\hat{S}}\hat{H}e^{-i\hat{S}} = e^{2i\hat{S}}\hat{H}$$

$$= \left\{\cos\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right) + \frac{\hat{\beta}(\hat{\alpha}\cdot\hat{\mathbf{p}})}{|\mathbf{p}|}\sin\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right)\right\}\{c(\hat{\alpha}\cdot\hat{\mathbf{p}}) + \hat{\beta}mc^2\}$$

$$= \frac{(\hat{\alpha}\cdot\hat{\mathbf{p}})}{|\mathbf{p}|}\left\{c|\hat{\mathbf{p}}|\cos\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right) - mc^2\sin\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right)\right\}$$

$$+\hat{\beta}\left\{mc^2\cos\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right) + c|\hat{\mathbf{p}}|\sin\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right)\right\}. \tag{10.87^2}$$

We want the term containing $\hat{\alpha}$ to vanish. The necessary condition is

$$\tan\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right) = \frac{|\hat{\mathbf{p}}|}{mc}, \text{ or, } \Theta = \frac{mc}{|\hat{\mathbf{p}}|}\tan^{-1}\frac{|\hat{\mathbf{p}}|}{mc}. \tag{10.93}$$

Then,

$$\cos\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right) = mc^2/\epsilon_{\mathbf{p}}\,;\ \sin\left(\frac{|\hat{\mathbf{p}}|\Theta}{mc}\right) = c|\hat{\mathbf{p}}|/\epsilon_{\mathbf{p}}. \tag{10.93a}$$

so that

$$\hat{H}' = \hat{\beta}\sqrt{\hat{\mathbf{p}}^2c^2 + m^2c^4} = \hat{\beta}\epsilon_{\mathbf{p}}.$$

[Note that $\hat{\mathbf{p}}' = e^{i\hat{S}}\hat{\mathbf{p}}e^{-i\hat{S}} = \hat{\mathbf{p}}$].

The transformation (10.90) was first proposed by Foldy and Wouthuysen.[21] Hence the name Foldy-Wouthuysen (F-W) transformation. The transformed wave function is given by

$$\psi' = \hat{U}\psi = e^{i\hat{S}}\psi, \tag{10.94}$$

so that the wavefunction corresponding to the positive energy state is

$$\psi'_+ = \frac{1}{2}(\hat{1} + \hat{\beta})\psi', \tag{10.95a}$$

Similarly, negative energy states are represented by

$$\psi'_- = \frac{1}{2}(\hat{1} - \hat{\beta})\psi'. \tag{10.95b}$$

---

21. Foldy L.L. and Wouthuysen S.A., *Phys. Rev.*, **78**, 29 (1950).

ψ' is referred to as the wavefunction in the Foldy-Wouthuysen *representation*. A constant of motion will be represented by an even operator in the F-W representation. Therefore, from (10.67) and (10.74), we have,

$$\langle \psi'_{b+} | \hat{\Omega}' | \psi'_{a+} \rangle = \langle \phi'_b | \hat{\Omega}'^{(+)} | \phi'_a \rangle,$$

and

$$\langle \psi'_{b-} | \hat{\Omega}' | \psi'_{a-} \rangle = \langle \chi'_b | \hat{\Omega}'^{(-)} | \chi'_a \rangle.$$

Thus, the F-W transformation achieves a separation between the positive and the negative energy solutions of the Dirac equation. The positive energy solutions are described entirely in terms of 2-component wavefunctions. From Eqs. (10.95a, b) we see that the projection operators $\hat{\pi}'_+$ and $\hat{\pi}'_-$ that project out from an arbitrary state the positive energy and the negative energy parts, respectively, are given by [cf. Eq. (10.62)],

$$e^{iS}\hat{\pi}_\pm e^{-iS} = \hat{\pi}'_\pm = \frac{1}{2}(\hat{1} \pm \hat{\beta}), \tag{10.96a}$$

where $\hat{\pi}_\pm$ are the projection operators in an arbitrary representation. Using Eqs. (10.92b), (10.87[2]) and (10.93a), we get,

$$\begin{aligned}
\hat{\pi}_\pm &= e^{-iS}\hat{\pi}' \pm e^{iS} \\
&= e^{-iS}\frac{1}{2}(\hat{1} \pm \hat{\beta})e^{+iS} \\
&= \frac{1}{2} \pm \frac{1}{2}\hat{\beta}e^{2iS} \\
&= \frac{1}{2}(\hat{1} \pm \hat{H}/\epsilon_p).
\end{aligned} \tag{10.96b}$$

In the presence of an external field, transitions between positive and negative energy states are inevitable. Therefore, a perfect separation between the positive and the negative energies is not feasible. However, it is possible to find a F-W transformation[22] which leads to a separation of the positive and negative energies to some desired order in (v^2/c^2). That is, the transformed Hamiltonian is given as a power series in (v^2/c^2) with the odd operators eliminated from terms of up to the chosen order in (v^2/c^2). This procedure is, obviously, useful only when the external field is weak and (v^2/c^2) is sufficiently small.

Because of the occurrence of the momentum operator (which is a differential operator in configuration space) in $\hat{S}$, the F-W transformation (10.94) is a *non-local* one. That is, it depends on the values of ψ at different points separated by a distance. Writing

$$e^{+iS} = \frac{1}{2}\left(\frac{2\epsilon_p}{\epsilon_p + mc^2}\right)^{1/2}\left[\hat{1} \pm \frac{c\hat{\beta}(\hat{\alpha}\cdot\hat{p}) \pm mc^2}{\epsilon_p}\right], \tag{10.91a}$$

2. See Ref. 1, Section 4.3.

which follows from (10.91) and (10.93a), and assuming $\epsilon_p \gg mc^2$, which is the case at ultra high velocities, we get,

$$\psi'(x) \approx [\hat{1} - i\hbar c(\hat{\beta}/\epsilon_p)(\hat{\alpha}\cdot\nabla)]\psi(x). \tag{10.94a}$$

Thus, the contributions to ψ' come from a region of linear dimension of the order of

$$\Delta x \sim |i\hbar c\hat{\beta}\hat{\alpha}|/\epsilon_p \leq (\hbar/mc),$$

since

$$\epsilon_p = mc^2(1 - v^2/c^2)^{-1/2} \geq mc^2.$$

This result suggests that the position co-ordinates of a Dirac particle also might be defined only to within a Compton wave length. Indeed difficulties are encountered when we try to interpret $\hat{\mathbf{r}}$ as the operator representing the position co-ordinates. For example, the equation,

$$\frac{d\hat{\mathbf{r}}}{dt} = \frac{1}{i\hbar}[\hat{\mathbf{r}}, \hat{H}] = c\hat{\alpha}, \tag{10.97^1}$$

should normally imply [see, Eq. (4.32a) which should hold good here since the Dirac equation is of the form (4.14)] that $c\hat{\alpha}$ is the velocity operator for the Dirac particle. That it is not really so (except in the nonrelativistic limit) is evident from the noncommuting nature of its components and from the fact that its eigenvalues are $\pm c$. In fact, since $\hat{\alpha}$ mixes the positive and negative energy states, eigenvectors of $c\hat{\alpha}$ (and, hence, also the eigenvalues) do not correspond to the physical states of a free Dirac particle. We can hope to get a sensible expression for the velocity of the particle only in a representation in which the positive and the negative energies are separated. Using (10.91a) and the relations, $\{(\hat{\alpha}\cdot\hat{\mathbf{p}}), \hat{\alpha}\} = 2\hat{\mathbf{p}}$ and $(\hat{\alpha}\cdot\hat{\mathbf{p}})^2 = \hat{\mathbf{p}}^2$, we obtain for the transform of $(c\hat{\alpha})$ in the F-W representation, the expression,

$$c\hat{\alpha}' \equiv e^{iS}(c\hat{\alpha})e^{-iS} = c\hat{\alpha} + \frac{c^2\hat{\mathbf{p}}\hat{\beta}}{\epsilon_p} - \frac{c^3\hat{\mathbf{p}}(\hat{\alpha}\cdot\hat{\mathbf{p}})}{\epsilon_p(\epsilon_p + mc^2)}. \tag{10.97^2}$$

The expectation value of this operator for a positive energy state is:

$$\langle\psi_+' | c\hat{\alpha}' | \psi_+'\rangle = c^2\hat{\mathbf{p}}/\epsilon_p, \tag{10.98}$$

which is the usual relativistic expression for the velocity of a particle. An operator that plays the role of the position co-ordinates of the particle also can be obtained in this case[23]. In the presence of external fields, however, the separation between positive and negative energies can be only approximate and a consistent quantum mechanical interpretation of the position and velocity operators for a Dirac particle becomes difficult.

Problem 10.9: Verify Eqs. (10.91a) and (10.96b).

Problem 10.10: Show that the transformation $\hat{U}'$ that eliminates the even operator $\hat{\beta}$ from $\hat{H}$ is given by

23. A more detailed discussion of this topic can be found in Ref. 3.

$$\hat{U}' = e^{i\hat{S}'} = \frac{1}{2\epsilon_p}\left[\epsilon_p + c\,|\hat{\mathbf{p}}| - mc^2\hat{\beta}\frac{(\hat{\boldsymbol{\alpha}}\cdot\hat{\mathbf{p}})}{|\mathbf{p}|}\right].$$

Also, show that this transformation separates the states belonging to different eigenvalues of the helicity operator

$$\hat{h} = (\hat{\boldsymbol{\sigma}}^D \cdot \hat{\mathbf{p}})/|\mathbf{p}|.$$

The Hydrogen Atom

As a final and conclusive test of the correctness of the Dirac equation as the relativistic wave equation of spin $\frac{1}{2}$ particles, we examine its predictions regarding the spectrum of the hydrogen atom[24]. The Hamiltonian for this case is,

$$\hat{H} = c(\hat{\boldsymbol{\alpha}}\cdot\hat{\mathbf{p}}) + \beta mc^2 + \hat{V}(r), \qquad (10.99^1)$$

with $\hat{V}(r) = -Ze^2/r$ [see Eq. (4.98a); for hydrogen atom, Z = 1]. We are interested in the stationary states of $\hat{H}$, which are solutions of the eigenvalue equation,

$$\hat{H}\psi = E\psi. \qquad (10.100^1)$$

As in the non-relativistic case, the angular momentum would be a constant of motion in a central field. Therefore, the equation (10.100^1) could be reduced to a radial equation by separating the radial and the angular-cum-spin parts of the wavefunction. The energy eigenvalues are then determined entirely by the radial equation. In the non-relativistic case, the separation is facilitated by the observation that the orbital angular momentum is a constant of motion or, in the presence of a spinorbit interaction, that $\hat{\mathbf{J}}^2, \hat{\mathbf{L}}^2, \hat{\mathbf{S}}^2, (\hat{\mathbf{S}}\cdot\hat{\mathbf{L}})$ and $\hat{J}_z$ are constants of motion. As a result, the wavefunction could be written as the product of a common eigenvector of $\hat{\mathbf{J}}^2, \hat{\mathbf{L}}^2, \hat{\mathbf{S}}^2$ and, since $(\hat{\mathbf{S}}\cdot\hat{\mathbf{L}}) = \frac{1}{2}(\hat{\mathbf{J}}^2 - \hat{\mathbf{L}}^2 - \hat{\mathbf{S}}^2)$, of $(\hat{\mathbf{S}}\cdot\hat{\mathbf{L}})$ and $\hat{J}_z$, and a function which depends only on the radial co-ordinates:

$$\Phi(\mathbf{r}) = R_l(r)\mathcal{Y}_m^{j(l,s)}. \qquad (10.101)$$

The fact that $\mathcal{Y}_m^{j(l,s)}$ is an eigenvector of $(\hat{\mathbf{S}}\cdot\hat{\mathbf{L}})$ ensures that $\Phi(\mathbf{r})$ is consistent with the relationship $\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}$.

In the relativistic case, the orbital angular momentum is not a constant of motion. $\hat{\mathbf{J}}^2, \hat{\mathbf{S}}^2, \hat{\mathbf{J}}_z$ and $\hat{\mathbf{L}}^2 + 2(\hat{\mathbf{S}}\cdot\hat{\mathbf{L}}) = \hat{\mathbf{J}}^2 - \hat{\mathbf{S}}^2$ commute with $\hat{H}$ so that we could write the wavefunction as the product of a radial function and a common eigenfunction of $\hat{\mathbf{J}}^2, \hat{\mathbf{S}}^2$ and $\hat{J}_z$. This procedure is, however, unsatisfactory because it does not incorporate the relationship $\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}}$. We should, therefore, find an operator $\hat{\mathcal{K}}$ which will play, in the relativistic case, the role of $(\hat{\mathbf{S}}\cdot\hat{\mathbf{L}})$ in the non-relativistic case. $\hat{\mathcal{K}}$ must, obviously, satisfy the following conditions:

(i) It should commute with $\hat{H}$.

(ii) It should commute with $\hat{\mathbf{J}}^2, \hat{\mathbf{S}}^2, \hat{J}_z$ and $\hat{\mathbf{L}}^2$.

---

24. The treatment given here largely follows that of Ref. 2.

Using these conditions, it is possible to obtain an expression for $\hat{\mathcal{K}}$ (Problem 10.11). The result is,

$$\hat{\mathcal{K}} = \hat{\beta}[(\hat{\sigma}^D \cdot \hat{\mathbf{L}}) + \hbar]. \tag{10.102}$$

Then,

$$\begin{aligned}\hat{\mathcal{K}}^2 &= \hat{\mathbf{L}}^2 + \hbar(\hat{\sigma}^D \cdot \hat{\mathbf{L}}) + \hbar^2 \\ &= \{\hat{\mathbf{L}} + (\hbar/2)\hat{\sigma}^D\}^2 + \frac{\hbar^2}{4} \\ &= \hat{\mathbf{J}}^2 + \frac{\hbar^2}{4}.\end{aligned} \tag{10.103}$$

Thus, the eigenvalues of $\hat{\mathcal{K}}^2$ are $\left\{\left(j+\frac{1}{2}\right)\hbar\right\}^2$. The eigenvalues of $\hat{\mathcal{K}}^2$ are, therefore, equal to $k\hbar$, where,

$$k = \pm\left(j+\tfrac{1}{2}\right) = \pm 1, \pm 2, \ldots, \pm\infty; \tag{10.104}$$

since, according to (5.49a) and (5.73b), j is a half-odd integer. We have two values of k for every value of j. This corresponds to the two values of $(\hat{\sigma}^D \cdot \hat{\mathbf{L}})$. We have[25]

$$\begin{aligned}\left\langle \frac{1}{\hbar}(\hat{\sigma}^D \cdot \hat{\mathbf{L}}) \right\rangle &= \frac{1}{\hbar^2}\left\langle \left[\hat{\mathbf{J}}^2 - \hat{\mathbf{L}}^2 - \frac{1}{4}\hat{\sigma}^{D2}\right] \right\rangle \\ &= j(j+1) - l(l+1) - \frac{3}{4} \\ &= -\left(j+\frac{3}{2}\right), \text{ for } l = j+\frac{1}{2} \\ &= \left(j-\frac{1}{2}\right), \text{ for } l = j-\frac{1}{2}\end{aligned}$$

Thus,

$$k = -\left(j+\frac{1}{2}\right), \text{ for } l = j+\frac{1}{2}, \tag{10.104a}$$

$$k = \left(j+\frac{1}{2}\right), \text{ for } l = j-\frac{1}{2}. \tag{10.104b}$$

We should now express $\hat{H}$ in terms of $\hat{\mathcal{K}}$ and operators that involve only radial variables. To this end, we define the Hermitian operators,

$$\hat{p}_r \equiv (-i\hbar\nabla)_r = -i\hbar\left(\frac{\partial}{\partial r} + \frac{1}{r}\right) = \frac{1}{r}(\hat{\mathbf{r}} \cdot \hat{\mathbf{p}} - i\hbar), \tag{10.105a}$$

---

25. Identifying a given k-value with a definite l-value [Eqs. (10.104a, b)] is strictly valid only in the non-relativistic limit as, in the relativistic case, l is not a good quantum number whereas k is.

$$\hat{\alpha}_r = \frac{1}{r}(\hat{\alpha}\cdot\hat{\mathbf{r}}),\ \text{or}\ (\hat{\alpha}\cdot\hat{\mathbf{r}}) = r\hat{\alpha}_r. \tag{10.105b}$$

We have,

$$\hat{\alpha}_r^2 = \hat{1};\ \{\beta, \hat{\alpha}_r\} = \hat{0};\ [\hat{\alpha}_r, \hat{\mathcal{K}}] = \hat{0}.$$

$$(\hat{\alpha}\cdot\hat{\mathbf{r}})(\hat{\alpha}\cdot\hat{\mathbf{p}}) = (\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}) + i(\hat{\sigma}^{(D)}\cdot\hat{\mathbf{L}}) = \hat{r}\hat{p}_r + i\hat{\beta}\hat{\mathcal{K}}$$

(by problem 10.3)

or,

$$(\hat{\alpha}\cdot\hat{\mathbf{p}}) = \hat{\alpha}_r\hat{p}_r + \frac{\hat{i}\alpha_r\hat{\beta}\hat{\mathcal{K}}}{r}. \tag{10.106}$$

Substituting from (10.106) in (10.99[1]):

$$\hat{H} = c\hat{\alpha}_r\hat{p}_r + \frac{ic\hat{\alpha}_r\hat{\beta}\hat{\mathcal{K}}}{r} + \hat{\beta}mc^2 + \hat{V}(r). \tag{10.99[2]}$$

We choose a representation in which $\hat{H}, \hat{J}_z$ and $\hat{\mathcal{K}}$ are diagonal (Since $\hat{\mathbf{J}}^2$ and $\mathbf{S}^2$ commute with $\hat{H}$, $\hat{\mathcal{K}}$and $\hat{J}_z$, they are also diagonal) and denote the eigenvectors by $\psi_{n,k,m}(x)$:

$$\hat{H}\psi_{n,k,m} = E_{n,k}\psi_{n,k,m},$$

$$\hat{\mathcal{K}}\psi_{n,k,m} = k\hbar\psi_{n,k,m},$$

$$\hat{J}_z\psi_{n,k,m} = m\hbar\psi_{n,k,m}. \tag{10.107}$$

(Just as in the non-relativistic case the energy depends on l, in the present case E should depend on k). The eigenvalue equation (10.100[1]) becomes,

$$[c\hat{\alpha}_r\hat{p}_r + i\hbar ckr^{-1}\hat{\alpha}_r\hat{\beta} + \hat{\beta}mc^2 + \hat{V}(r) - E_{n,k}]\psi_{n,k,m} = 0. \tag{10.100[2]}$$

ψ cannot still be written as the product of a radial part and an angular-cum-spin part because of the occurrence in (10.100[2]) of $\hat{\alpha}_r$ and $\hat{\beta}$ which do not commute with $\hat{H}$. In order to eliminate these operators, we left-multiply (10.100[2]) by $\hat{\beta}$, add the result to (10.100[2]) and then divide by two, getting,

$$\left(c\hat{p}_r - \frac{i\hbar ck}{r}\right)(\hat{\alpha}_r\psi_{-n,k,m}) + (mc^2 + \hat{V}(r) - E_{n,k})\psi_{+n,k,m} = 0 \tag{10.108[1]}$$

where,

$$\psi_{\pm n,k,m} = \tfrac{1}{2}(\hat{1}\pm\hat{\beta})\psi_{n,k,m}.$$

Similarly, $\tfrac{1}{2}[\hat{\alpha}_r\times(10.100^2) + \hat{\beta}\hat{\alpha}_r\times(10.100^2)]$ gives,

$$\left(c\hat{p}_r + \frac{i\hbar ck}{r}\right)\psi_{+n,k,m} - [mc^2 + E_{n,k} - \hat{V}(r)]\,(\hat{\alpha}_r\psi_{-n,k,m}) = 0. \tag{10.109[1]}$$

Eqs. (10.108[1]) and (10.109[1]) constitute a set of coupled equations for $\psi_{+n,k,m}$ and $\alpha_r\psi_{-n,k,m}$. Since the coefficients do not contain either spin or angle-dependent operators, ψ_+ and $\alpha_r\psi_-$ could be written as products of a radial function and a spinor. The latter has to be the same for both ψ_+ and $\hat{\alpha}_r\psi_-$ as it is determined by $\hat{\mathcal{K}}$only:

$$\underset{+n,k,m}{\psi(x)} = \phi_{n,k}(r)\xi_{k,m}, \tag{10.110a}$$

$$\underset{-n,k,m}{\hat{\alpha}_r\psi(x)} = \chi_{n,k}(r)\xi_{k,m} \tag{10.110b}$$

Substituting these in (10.108[1]) and (10.109[1]) and defining,

$$\Delta_1 = \frac{mc^2 + E_{n,k}}{\hbar c}, \Delta_2 = \frac{mc^2 - E_{n,k}}{\hbar c},$$

$$\Delta = +\sqrt{\Delta_1\Delta_2} = (i/\hbar)p_{n,k}, \tag{10.111}$$

$$\Gamma = (Ze^2/\hbar c) = -rV(r)/\hbar c,$$

$$\rho = \Delta r, \frac{F(r)}{r} = \phi_{n,k}, \frac{G(r)}{r} = i\chi_{n,k}$$

we get,

$$\left(\frac{d}{d\rho} + \frac{k}{\rho}\right)G(\rho) - \left(\frac{\Delta_2}{\Delta} - \frac{\Gamma}{\rho}\right)F(\rho) = 0, \tag{10.108^2}$$

$$\left(\frac{d}{d\rho} - \frac{k}{\rho}\right)F(\rho) - \left(\frac{\Delta_1}{\Delta} + \frac{\Gamma}{\rho}\right)G(\rho) = 0. \tag{10.109^2}$$

These equations can be solved by the method of series integration. The wave functions ϕ and χ should be finite at the origin $r = 0$ but should vanish at infinity. Thus,

$$G(\rho = 0) = 0 = F(\rho = 0), \tag{10.112}$$

$$G(\rho = \infty) = 0 = F(\rho = \infty),$$

which suggests the expansions,

$$F(\rho) = f(\rho)e^{-\rho}; G(\rho) = g(\rho)e^{-\rho}, \tag{10.113}$$

where,

$$f(\rho) = \rho^{\nu}(f_0 + f_1\rho + f_2\rho^2 + \ldots + f_{n'}\rho^{n'})$$

$$g(\rho) = \rho^{\nu}(g_0 + g_1\rho + g_2\rho^2 + \ldots + g_{n'}\rho^{n'}), \tag{10.114}$$

with

$$f_0 \neq 0, g_0 \neq 0, \tag{10.114a}$$

and

$$n' = 0, 1, 2, \ldots, \text{(any positive integer)} \tag{10.114b}$$

The series (10.114) are finite as otherwise the boundary conditions at $\rho = \infty$ cannot be satisfied. Also, the boundary conditions at $\rho = 0$ require

$$f(0) = g(0) = 0,$$

so that, $$\nu > 0. \tag{10.115a}$$

Substituting from (10.113) and (10.108^2) into (10.109^2), we get,

$$\left(\frac{d}{d\rho}-1+\frac{k}{\rho}\right)g-\left(\frac{\Delta_2}{\Delta}-\frac{\Gamma}{\rho}\right)f=0, \tag{10.108^3}$$

$$\left(\frac{d}{d\rho}-1-\frac{k}{\rho}\right)f-\left(\frac{\Delta_1}{\Delta}+\frac{\Gamma}{\rho}\right)g=0. \tag{10.109^3}$$

Substituting for f and g from (10.114) and equating the coefficients of the various powers of ρ to zero, we have:

$$\rho^{\nu-1}: (\nu+k)g_0+\Gamma f_0=0, \tag{10.116^1}$$

$$-\Gamma g_0+(\nu-k)f_0=0.$$

Since $f_0\neq 0$ and $g_0\neq 0$, this requires

$$\nu=\sqrt{k^2-\Gamma^2}<|k|. \tag{10.115b}$$

$$\left.\begin{aligned} p^{\nu+s-1}: (\nu+s+k)g_s-g_{s-1}+\Gamma f_s-\frac{\Delta_2}{\Delta}f_{s-1}=0,\\ (\nu+s-k)f_s-f_{s-1}-\Gamma g_s-\frac{\Delta_1}{\Delta}g_{s-1}=0 \end{aligned}\right\} \tag{10.117^1}$$

Eliminating g_{s-1} and f_{s-1} from these equations, we get,

$$\frac{g_s}{f_s}=\frac{(\nu+s-k)\Delta_2-\Gamma\Delta}{\Gamma\Delta_2+(\nu+s+k)}, \tag{10.118^1}$$

From (10.116^1),

$$\frac{g_0}{f_0}=\frac{\nu-k}{\Gamma}=-\frac{\Gamma}{\nu+k}. \tag{10.116^2}$$

Putting $s=n'+1$ in (10.117^1), we have,

$$g_{n'}=-(\Delta_2/\Delta)f_{n'}, \tag{10.117^2}$$

while from (10.118^1), we get,

$$g_{n'}=\left\{\frac{(\nu+n'-k)\,\Delta_2-\Gamma\Delta}{\Gamma\Delta_2+(\nu+n'+k)\Delta}\right\}f_{n'}. \tag{10.118^2}$$

From (10.117^2) and (10.118^2), we derive, using (10.111),

$$2\Delta(\nu+n')=\Gamma(\Delta_1-\Delta_2)=(2\Gamma/\hbar c)E_{n',k}. \tag{10.119}$$

Squaring this expression leads to:

$$E_{n'k}=mc^2\left[1+\frac{\Gamma^2}{(\nu+n')^2}\right]^{-1/2}, (n'=0,1,2,\ldots) \tag{10.120}$$

where the positive character of $E_{n',k}$ evident from (10.119), is taken into account.

Formula (10.120) gives the energy levels of the hydrogen atom. Its predictions (after taking into account the hyperfine splitting of each level into a doublet due to the interaction of the proton spin with the electron magnetic moment) are in agreement with the observed spectrum of hydrogen, except for minor discrepancies discovered much later (1947) and known as *Lamb shift*.

The *fine structure* of the levels can be exhibited by expanding (10.120) in powers of Γ^2, remembering[26]

$$(\nu+n')^2=(\sqrt{k^2-\Gamma^2}+n')^2$$

$$\approx n^2\left[1-\left\{\frac{\Gamma^2}{n^2}\left(1+\frac{n'|k|}{k^2}\right)+\frac{\Gamma^4}{n^2}\cdot\frac{n'|k|}{4k^4}\ldots\right\}\right].$$

where, $$n=n'+|k|. \tag{10.121}$$

The result is, up to second order in Γ^2,

$$E_{n,k}=mc^2\left[1-\frac{\Gamma^2}{2n^2}-\frac{\Gamma^4}{2n^4}\left(\frac{n}{|k|}-\frac{3}{4}\right)\right]. \tag{10.120a}$$

n is called the *total* quantum number and n' the *radial* quantum number. The range of values of (n, n') and k are given by Eq. (10.121) when the allowed values of k and n' [Eqs. (10.104) and (10.114b)] are taken into account. We have,

$$\begin{aligned} n&=1,2,3,\ldots,+\infty,\\ n'&=0,1,2,\ldots,(n-1),\\ k&=\pm(n-n'),\ \text{when } n'\neq 0\\ &=(n-n'),\quad \text{when } n'=0. \end{aligned} \tag{10.122}$$

The last line in (10.122) is deduced as follows: When $n'=0$, we have two different expressions for the ratio (g_0/f_0). From (10.117[2]), we have, $(g_0/f_0)=-(\Delta_2/\Delta)<0$. But from (10.116[2]), $(g_0/f_0)=(\nu-k)/\Gamma$. Since $0<\nu<|k|$ and $\Gamma>0$, $(\nu-k)/\Gamma$ is negative when k is positive and positive when k is negative. Consistency of the two expressions for the ratio thus requires that k be positive.

Eq. (10.120a) shows that the level corresponding to a given n, splits up into n levels corresponding to the n different values of $|k|$. These levels constitute the *fine structure* of the level labelled by n. This name derives from the fact that the energy separation of the levels corresponding to different $|k|$ but the same n is small compared with the energy separation of levels belonging to different n. In fact, the ratio of the former to the latter is seen from (10.120a) to be of the order of $\Gamma^2\approx(1/137)^2$ for the hydrogen atom. We also see that the energy levels are two-fold degenerate with respect to the k-value since $E_{n,-k}=E_{n,k}$ (This degeneracy is removed in the 'Lamp shift'). The fine structure belonging to $n=3$ is displayed in Table 10.1, where the levels are classified by the quantum numbers n', k and $j=|k|-\frac{1}{2}$. It is also customary to label the states by the non-relativistic quantum number l (related to k by Eqs. (10.104 a, b)) and the spectroscopic notation based on that. These labels are also given in the Table where, for example $3P_{3/2}$ denotes the state with $n=3, j=3/2$ and $l=1$ $(k=2)$. In Fig. 10.2, a (schematic) comparison

26. Γ is known as the *fine structure constant*. For hydrogen atom

$$\Gamma=(e^2/\hbar c)\approx 1/137.$$

of the observed and the predicted (according to formula (10.120a)) levels of hydrogen for $n = 3$ is presented.

Table 10.1. Fine structure of energy levels of hydrogen atom for n = 3

Relativistic labels				Non-relativistic labels	
n'	k	j	l	Spectroscopic label	
0	3	5/2	2	$3D_{5/2}$	spin-orbit splitting
1	−2	3/2	2	$3D_{3/2}$	degenerate
	2	3/2	1	$3P_{3/2}$	
2	−1	1/2	1	$3P_{1/2}$	
	1	1/2	0	$3S_{1/2}$	

Problem 10.11: Assuming $\hat{\mathcal{K}}$ to be of the form $\hat{\mathcal{K}} = \hat{A}(\hat{\sigma}^D \cdot \hat{\mathbf{L}}) + \hat{B}$, derive the result (10.102).

Problem 10.12: Calculate the energies of the hydrogen levels up to $n = 3$, using formula (10.120a). Plot these on an energy level diagram similar to that of Fig. 10.2. [$mc^2 = 510$ keV; $\Gamma = 1/137$].

Lorentz Covariance of the Dirac Equation

Now that sufficient evidence is at our disposal to suggest that the Dirac equation is the relativistic quantum mechanical wave equation of the electron, we should turn our attention to establishing its covariance under Lorentz transformations. The form (10.25^2) has the appearance of a covariant equation. Nevertheless it is necessary to establish the Lorentz covariance explicitly; for (as will be clear from what follows) covariance cannot be taken for granted merely on the basis of the appearance of the equation.

Now, Lorentz covariance of an equation means that it should have the same *form* in all inertial frames. Thus, if

$$(\gamma_\mu \partial_\mu + \kappa)\psi(x) = 0, \qquad (10.25^2)$$

is the Dirac equation in the inertial frame S, then it should be of the form,

$$(\gamma_\mu \partial_\mu' + \kappa)\psi'(x') = 0, \qquad (10.123)$$

in the inertial frame S', where ∂_μ', x' and $\psi'(x')$ are, respectively, the Lorentz transforms of ∂_μ, x and $\psi(x)$. A Lorentz transformation is an orthogonal transformation in the Minkowski space and could, therefore, be represented by the co-ordinate transformation,

$$x_\mu \to x_\mu' = a_{\mu\nu} x_\nu, \qquad (10.124)$$

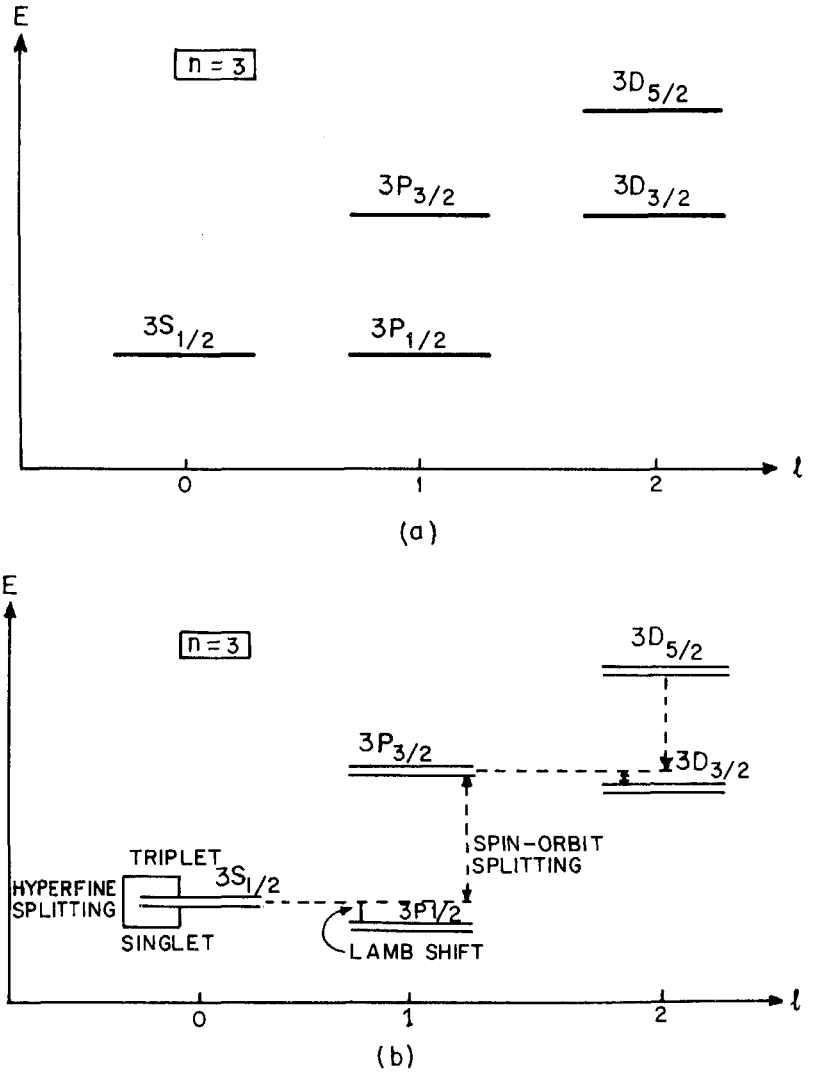


Fig. 10.2. Energy levels of the hydrogen atom belonging to $n = 3$, (a) Dirac theory, (b) observed.

or,

$$x' = ax.$$

The orthogonality of the transformation is denoted by the condition that 'a' be an orthogonal matrix. That is [see Eq. (A. 39)],

$$\tilde{a} = a^{-1},$$

or,

$$a_{\mu\nu}a_{\mu\lambda} = a_{\nu\mu}a_{\lambda\mu} = \delta_{\nu\lambda}, \tag{10.125a}$$

Also,

$$\det a = \pm 1. \tag{10.125b}$$

The transformations with $\det a = +1$ are called *proper* Lorentz transformations while those with $\det a = -1$ are called *improper* Lorentz transformations. The former include rotations in three dimensions in addition to co-ordinate transformation corresponding to uniform relative motion along a spatial direction, whereas the latter refers to the discrete transformations of space inversion and time reversal (see sections 6.2D and 6.2E).

From (10.124), we have,

$$\partial_\mu = \frac{\partial x_\nu'}{\partial x_\mu}\partial_\nu' = a_{\nu\mu}\partial_\nu'. \tag{10.126}$$

Substituting from (10.126) in (10.25[2]), we get,

$$(\bar{\gamma}_\nu \partial\, \mu' + \kappa)\, \psi(x) = 0, \tag{10.127[1]}$$

where,

$$\bar{\gamma}_\nu = a_{\nu\mu}\gamma_\mu. \tag{10.128[1]}$$

The result,

$$\{\bar{\gamma}_\mu, \bar{\gamma}_\nu\} = 2\delta_{\mu\nu}, \tag{10.129}$$

is easily verified. Thus, the $\bar{\gamma}_\mu$'s obey the same algebra (Eq. (10.27)) as the γ_μ's. But we have seen that the algebra (10.27) has only one irreducible representation. Therefore $\bar{\gamma}_\nu$ should be related to γ_ν through a similarity transformation:

$$\bar{\gamma}_\nu \equiv a_{\nu\mu}\gamma_\mu = L^{-1}(a)\gamma_\nu L(a), \tag{10.128[2]}$$

L being a nonsingular matrix. Eq. (10.127[1]) then becomes,

$$(\gamma_\nu \partial_{\nu'} + \kappa)L\ \psi(x) = 0. \tag{10.127[2]}$$

Comparing Eq. (10.127[2]) with (10.123), we see that

$$\psi'(x') \equiv \psi'(ax) = L(a)\psi(x), \tag{10.130}$$

where L is defined by Eq. (10.128[2]). What we have to show to establish the Lorentz covariance of the Dirac equation is the *existence* of a matrix L satisfying Eq. (10.128[2]) corresponding to every Lorentz transformation defined by the orthogonal matrix 'a'. Eq. (10.130) then gives the law of transformation of the Dirac wavefunction.

Case 1. Proper Lorentz Transformations

A proper Lorentz transformation is a continuous transformation and, as such, can be built up from a succession of infinitesimal transformations. It is, therefore, sufficient to establish the existence of an L satisfying Eq. (10.128[2]) in the case of an infinitesimal proper Lorentz transformation.

Now, an infinitesimal Lorentz transformation is defined by,

$$a_{\mu\nu} = \delta_{\mu\nu} + \epsilon_{\mu\nu}, \; |\epsilon_{\mu\nu}| \ll 1. \tag{10.131}$$

Then, from (10.125a), we have,

$$\epsilon_{\mu\nu} = -\epsilon_{\nu\mu}, \tag{10.131a}$$

$$\epsilon_{\mu\mu} = 0.$$

The transformation is, thus, described in terms of six parameters. Substituting from (10.131) in (10.128[2]), we get,

$$\gamma_\nu + \epsilon_{\nu\mu}\gamma_\mu = L^{-1}(\epsilon)\,\gamma_\nu L(\epsilon) \tag{10.128[3]}$$

This equation is satisfied to first order in $\epsilon_{\nu\mu}$ by the choice,

$$L = I + \frac{1}{4}\epsilon_{\mu\nu}\gamma_\mu\gamma_\nu. \tag{10.132}$$

This gives,

$$L^{-1} = I - \frac{1}{4}\epsilon_{\mu\nu}\gamma_\mu\gamma_\nu. \tag{10.133}$$

Since L depends only on the Dirac matrices (aside from the parameters defining the Lorentz transformation), it follows that we can always find an L corresponding to a given infinitesimal proper Lorentz transformation, and, hence, corresponding to any proper Lorentz transformation.

As an example, let us consider an infinitesimal rotation of the reference frame through angle $\delta\phi$ about the z-axis. In this case, we have (cf. Fig. 5.8),

$$x_1' = x_1 + \delta\phi x_2,$$

$$x_2' = x_2 - \delta\phi x_1,$$

$$x_3' = x_3 \quad ,$$

$$x_4' = x_4 \quad ,$$

so that the matrix ϵ is given by

$$\epsilon = \begin{pmatrix} 0 & \delta\phi & 0 & 0 \\ -\delta\phi & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus,

$$L(\mathrm{e}_3, \delta\phi) = I + \frac{i}{4}\epsilon_{\mu\nu}\sigma_{\mu\nu}$$

$$= I + (i/2)\delta\phi\,\sigma_{12},$$

where,

$$\sigma_{\nu\mu} = i\gamma_\mu\gamma_\nu = -\sigma_{\mu\nu}, \tag{10.134}$$

and $\mathbf{e}_3$ is a unit vector in the z-direction.

From Eqs. (10.36), (10.38a) and (10.134), we see that

$$\sigma_{12} = \sigma_3^D.$$

Then,

$$L(\mathbf{e}_3, \delta\phi) = I + (i/2)\delta\phi\,\sigma_3^D \qquad (10.135^1)$$

In general, for an infinitesimal rotation about an axis $\mathbf{n}$, we have,

$$L(\mathbf{n}, \delta\phi) = I + (i/2)\,\delta\phi\,(\mathbf{n}\cdot\sigma^D), \qquad (10.135^2)$$

from which we see that the matrix corresponding to a finite rotation ϕ is [see Eq. (5.135)],

$$L(\mathbf{n}, \phi) = \exp[(i/2)\,\phi\,(\mathbf{n}\cdot\sigma^D)]. \qquad (10.135^3)$$

This equation represents the peculiar transformation law applicable to a spinor: a rotation through 4π is required to return $\psi(x)$ to its original value.

The matrix L corresponding to spatial rotations is unitary [as seen from (10.135)]. However, in general L is not unitary (as for example, in the case of a uniform relative motion and in the case of time reversal).

Problem 10.13: Obtain the matrix L corresponding to uniform relative motion with velocity v along the z-axis.

Case 2: Improper Lorentz Transformations

These include space inversion and time reversal.

Space Inversion

This is defined by

$$x_k \to x_k' = -x_k;\; x_4' = x_4'\,(k = 1, 2, 3), \qquad (10.136a)$$

so that,

$$a_{ik} = -\delta_{ik};\; a_{4k} = a_{k_4} = 0;\; a_{44} = 1. \qquad (10.136b)$$

Substituting from (10.136b) in (10.128^2), we get,

$$L\gamma_k + \gamma_k L = 0,\; (k = 1, 2, 3,), \qquad (10.128^4)$$

$$L\gamma_4 - \gamma_4 L = 0.$$

Thus, L commutes with γ_4 and anticommutes with γ_k. It follows that

$$L\ (\text{space inversion}) = (e^{i\xi})\gamma_4. \qquad (10.137)$$

L is a representation of the parity operator $\hat{P}$. From Eq. (10.130), we have,

$$\hat{P}\,\psi(x) = e^{i\xi}\,\hat{\gamma}_4\psi(x). \qquad (10.130a)$$

In the nonrelativistic limit, $\hat{\gamma}_4 \equiv \hat{\beta} \to \pm\hat{1}$ [Eq. (10.79^1)], so that the positive and the negative energy states become eigenstates of $\hat{P}$ (ξ could be chosen to be zero in this case) with opposite eigenvalues (that is, *intrinsic parities*).

The invariance of the Dirac theory under the parity operation follows from the invariance of the Dirac Hamiltonian (10.21^2) [both $\hat{\alpha}$ and $\hat{\mathbf{p}}$ change sign by Eqs. (10.128^4) and (6.13b)].

Time Reversal

We do not wish to give a detailed treatment of this transformation here as it has already been discussed in Section 6.2E. It is the transformation in which all directions of motion are reversed (or, $x_4 \to -x_4$) and, as we have seen in Section 6.2E, it is antilinear. Because of this antilinearity, Eq. (10.128^2) needs a slight modification. For, in obtaining Eq. (10.127^2) from (10.127^1) we have used the relation, $L\partial_\nu'\psi = \partial_\nu' L\psi$. But when L represents an antilinear transformation, $L\partial_4'\psi = -\partial_4' L\psi$. Then, (10.127^1) leads to (10.127^2) only if

$$a_{k\mu}\gamma_\mu = L^{-1}\gamma_k L, (k = 1,2,3)$$

and

$$a_{4\mu}\gamma_\mu = -L^{-1}\gamma_4 L. \qquad (10.128^5)$$

Since

$$a = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

We see that L should commute with all four γ_μ. An explicit expression for L can be obtained by a generalisation of the nonrelativistic case:

According to Problem 6.6, the non-relativistic time-reversal operator $\hat{T}$ for the case of a spin $\frac{1}{2}$ particle is given by,

$$\hat{T} = i\,\hat{\sigma}_2 \hat{K}_c, \qquad (10.138^1)$$

where, $\hat{K}_c$ is the operator corresponding to complex conjugation and $\hat{\sigma}_2$ represents the 2-component of the Pauli spin vector. We expect to get the relativistic time-reversal operator simply by replacing $\hat{\sigma}_2$ by the 2-component of the Dirac spin vector. Thus,

$$L \text{ (time reversal)} \equiv \hat{T} = i\hat{\sigma}_2^D \hat{K}_c = \hat{\gamma}_3\hat{\gamma}_1\hat{K}_c, \qquad (10.138^2)$$

where expression (10.38a) for $\hat{\sigma}_2^D$ is used.

Bilinear Covariants

By a procedure similar to the one leading to Eq. (10.127^2), we find that the adjoint Dirac equation is transformed, under a Lorentz transformation, into

$$\partial_\nu' \bar{\psi}(x) L^{-1} \gamma_\nu - \kappa \bar{\psi}(x) L^{-1} = 0 \tag{10.139}$$

Comparing this with,

$$\partial_\nu' \bar{\psi}'(x') \gamma_\nu - \kappa \bar{\psi}'(x') = 0, \tag{10.140}$$

which is the adjoint Dirac equation in the frame S', we see that the Lorentz transform of $\bar{\psi}(x)$ is given by

$$\bar{\psi}'(x') = \bar{\psi}(x) L^{-1}. \tag{10.141}$$

Physical observables in the Dirac theory are bilinear in $\bar{\psi}$ and ψ [for example, j_μ given by (10.50)]. In fact, they are of the form $(\bar{\psi} \Gamma^n \psi)$ where Γ^n is one of the sixteen linearly independent 4×4 matrices that can be constructed out of the four Dirac matrices. The combination $(\bar{\psi} \Gamma^n \psi)$ is called a *bilinear covariant* because it transforms covariantly (that is, as the component of a 4-tensor). The sixteen matrices are (as already listed under properties of the Dirac matrices):

$$\Gamma^S = I;\ \Gamma_\mu^V = i\gamma_\mu;\ \Gamma_{\mu\nu}^T = \sigma_{\mu\nu} = -i\gamma_\mu\gamma_\nu;$$

$$\Gamma_\mu^A = i\gamma_5\gamma_\mu;\ \Gamma^P = i\gamma_5, \tag{10.142}$$

where, $\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4 = \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix}$. (10.143)

γ_5 satisfies, $\gamma_5\gamma_\mu = -\gamma_\mu\gamma_5;\ \gamma_5^2 = I$. (10.143a)

The following properties of the Γ^n are easily verified:

(1) $(\Gamma^n)^2 = \pm I$.

(2) Except for Γ^S, $\Gamma^n\Gamma^m = -\Gamma^m\Gamma^n$, $(n \neq m)$, so that $Tr(\Gamma^n) = 0, Tr(\Gamma^S) = 4$.

(3) $\Gamma^m\Gamma^l = c\Gamma^n$, $(n \neq S, m \neq l)$
where
$$c = \pm 1, \pm i.$$
From the above properties, it follows that the Γ^n are linearly independent. For, let $\sum_n a_n\Gamma^n = 0$. Then, multiplying by Γ^m and then taking the trace, we have,
$$a_m Tr[(\Gamma^m)^2] = \pm 4a_m = 0.$$

(4) An arbitrary 4×4 matrix Ω in the space of these 16 matrices is given by
$$\Omega = \sum_n \omega_n \Gamma^n,$$
with
$$\omega_n = \frac{1}{4} Tr(\Omega\Gamma^n).$$

The bilinear forms $(\bar{\psi} \Gamma^n \psi)$ can be classified as follows on the basis of their transformation properties under proper and improper Lorentz transformations. The following relationships are helpful in deriving these.

$$L^{-1}\gamma_\mu L = a_{\mu\nu}\gamma_\nu, \text{ for proper L.T.}$$

$$= -\gamma_\mu\ (\mu = 1,2,3) \text{ and } +\gamma_\mu\ (\mu = 4), \text{ for space inversion}$$

$$L^{-1}\gamma_5 L = \gamma_5, \text{ for proper L.T.}$$

$$= -\gamma_5, \text{ for space inversion.}$$

We find:

$S = \bar{\psi}\Gamma^S\psi$ – *a scalar* (invariant under both proper and improper L.T.)

$V_\mu = i\bar{\psi}\gamma_\mu\psi$ –*a vector* ($V_\mu' = a_{\mu\nu}V_\nu$, for proper L.T.

$V_k' = -V_k$ and $V_4' = V_4$ under space inversion)

$T_{\mu\nu} = -i\ \bar{\psi}\gamma_\mu\gamma_\nu\psi$ – second rank, antisymmetric *tensor*

($T'_{\mu\nu} = a_{\mu\sigma}a_{\nu\rho}T_{\sigma\rho}$, for proper L.T.,

$T_{ik}' = T_{ik}$; $T'_{44} = T_{44}$; $T_i' = -T_i$, under space inversion)

$A_\mu = i\ \bar{\psi}\gamma_5\gamma_\mu\psi$ – *pseudo* or *axial vector*; $A_\mu' = a_{\mu\nu}A_\nu$ (pr. L.T.);

$A_k' = A_k$; $A_4' = -A_4$ (Impr. L.T.)

$P = i\ \bar{\psi}\gamma_5\psi$ – *pseudo scalar* (invariant under pr. L.T. but changes sign under space inversion).

The Hole Theory

By now, we have before us an impressive array of successes of the Dirac theory — the spin and the magnetic moments of the electron, the fine structure of the hydrogen atom, a positive definite probability density, invariance under both proper and improper Lorentz transformations, and so on. There is only one aspect of the theory that dampens our sense of triumph; the negative energy states. So far we have avoided facing the problem. But clearly a solution to the problem of the negative energy states has to be found if the Dirac theory is to survive. For, we have seen that the negative energy solutions cannot be lightly brushed aside since the positive energy solutions alone do not form a closed Hilbert space. As a result, a positive energy electron would be unstable against transition to negative energy states as soon as an interaction is switched on. In fact, the time it takes for such a transition can be shown to be extremely small.

Of course, the negative energy states would not have been a problem if negative energy electrons were actually found in nature. Such electrons would be accelerated opposite to the direction of an applied force. No such particles were observed.

It is against the above background that Dirac proposed[7], two years after the publication of his original theory, an *interpretation* known as a *hole theory*. The

objective was to explain why the negative energy electrons are not observed (though they exist). For this, Dirac relied on Pauli's Exclusion Principle (Section 9.2) according to which each quantum state of a spin $\frac{1}{2}$ particle can be occupied by only one such particle. He then proposed the following hypotheses which form the basis of the hole theory:

(H1) In the normal state, called the *vacuum*, all the negative energy states are occupied (in accordance with the exclusion principle) while all the positive energy states are empty.

(H2) Only deviations from the normal vacuum, and not the vacuum itself, can be observed.

Suppose we add a positive energy electron to the vacuum. This electron cannot make a transition to a negative energy state as all the negative energy states are already occupied. The stability of the positive energy electron is thus ensured.

Though the vacuum is unobservable, it is not inert; it can interact, by virtue of its electric charge, with an external electromagnetic field. Imagine, then, a gamma ray of energy $h\nu > 2mc^2$ interacting with the vacuum. One of the negative energy electrons in the vacuum can absorb the γ-ray and jump to a positive energy state, leaving a vacancy (a *hole*) behind [See Fig. 10.3(a)]; the minimum energy separating a positive energy state from a negative energy state is $mc^2 - (-mc^2) = 2mc^2$. How does the hole appear to an observer? If Q_0 and M_0 are the charge and mass, respectively, of the vacuum without the hole, and Q and M the same with the hole, then,

$$Q = Q_0 - e > Q_0, \text{ since } e < 0.$$

$$M = M_0 - (-m) = M_0 + m > M_0.$$

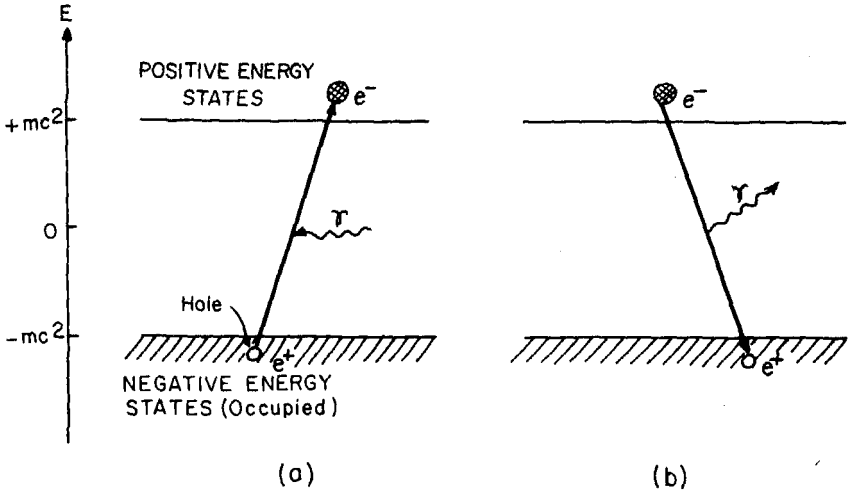


Fig. 10.3. Pair creation and annihilation in the Hole Theory.

Thus, the changes in the charge and mass of the vacuum are:

$$\Delta Q = Q - Q_0 = -e,$$

$$\Delta M = M - M_0 = +m.$$

That is, relative to the vacuum, the hole has the same mass as a positive energy electron but has a charge that is equal in magnitude but opposite in sign to that of the electron. In other words, the hole appears as a positively charged electron, which we will denote by the symbol e^+ (A normal electron is denoted by e^-). In addition, we have the electron which is promoted to the positive energy state. The whole process, thus, consists in the disappearance of a quantum of electromagnetic radiation and the appearance of a pair of oppositely charged electrons [Fig. 10.3(a)]. This phenomenon is called *pair creation.*

However, the newly created hole in the vacuum will not last long as an electron from the positive energy state will soon fall into it, giving up the excess energy ($\geq 2mc^2$) in the form of a quantum of electromagnetic radiation. In the process, both the electron and the hole disappear [Fig. 10.3(b)]. This phenomenon is thus the reverse of pair creation and is known as *pair annihilation.*

The hole theory has been beautifully vindicated by the later discovery[27] of the positively charged electron (named the *positron*) and the phenomena of pair creation and annihilation. The theory has been further supported by experimental verification, during the early 1950s, of another of its predictions, namely *vacuum polarization.* A positive energy electron repels the negative energy electrons (of the vacuum) in its neighbourhood. As a result, in the neighbourhood of the positive energy electron there in as excess of positive charge (less of negative charge) relative to the normal vacuum; that is, the vacuum is polarized. The *physical charge* e (the charge seen by an observer at a distance $> \hbar/mc$) of the electron includes this polarization charge in addition to its *bare charge* e_0. But an observer, or a charge (as in the case of the proton in a hydrogen atom in the S-state), close enough ($r \leq \hbar/mc$) to the electron sees a charge e', where $|e| < |e'| \leq |e_0|$. This effect leads to a lowering in energy of the S-states relative to the $l \neq 0$ states in the hydrogen atom [Eq. (10.120a)]. For the ground state, the shift amounts to about 10^{-2} eV. The shift appears as a correction to the Lamb shift which is much larger in magnitude and is in the opposite direction. The observed shift for the hydrogen levels is in agreement with the theoretical value for the Lamb shift[28] corrected for the effect of vacuum polarization.

The positron is not just another positively charged particle; for, it annihilates an electron on contact, itself getting annihilated by the electron in the process. It is the first (to come to light) of a new kind of particles called *antiparticles* (positron being the antielectron). It is believed[29] that all particles, including bosons and electrically neutral ones like the neutron and the neutrino, have their antiparticles[30].

---

27. Anderson, C.D. *Phys. Rev.* **41**, 405 (1932).
28. Lamb shift arises from the *self-energy* of the electron (that is, energy due to the interaction of an electron with its own field). The calculations are done within the framework of quantum field theory (specifically, quantum electro-dynamics).
29. The belief was reinforced by the discovery of the antiproton in 1956.
30. Some neutral particles are their own antiparticles. Examples: neutral pion, photon.

Thus, the hole theory not only solves the problem of the negative energy electrons, but also makes several important predictions which are eventually verified. Nevertheless, we cannot overlook certain profound implications of the theory as far as the relativistic wave equations in general and the Dirac equation in particular are concerned. First of all, it constitutes a renouncement of the original motivation that led to the rejection of the Klein-Gordon equation and the development of the Dirac equation. For, with the hole theory the Dirac equation no longer describes a system with a conserved number of particles (relative to the vacuum); in fact, it describes particles with different signs of the electric charge. It follows that it should be possible to rehabilitate the second order Klein-Gordon equation also with a suitable reinterpretation of the wavefunction. Secondly, the hole theory introduces an element of inconsistency in the Dirac theory. For, there is no way of incorporating into the Dirac equation, viewed as a quantum mechanical wave equation, phenomena such as pair creation and annihilation. These phenomena could be an integral part of only such a theory that does not have particle-number conservation (and, therefore, a positive-definite position probability density) as one of its basic features. Such a theory, namely the quantum field theory, was developed in later years. It would not be out of place to regard the hole theory as the harbinger of the quantum field theory.

10.2B. The Weyl Equations

As already stated, the wave equations (10.17a, b), derived for *zero-mass* particles, are known as the Weyl equations. The theory based on these wave equations is usually referred to as the *two-component theory of the neutrino*[31]. This name distinguishes it from another possible theory of the zero-mass spin $\frac{1}{2}$ particles, namely a 4-component theory based on the Dirac equation with $m = 0$. The basic difference between the two theories is that, whereas the Dirac theory is invariant under the parity operation (as we have already seen) the two-component theory is not. This follows from the expression $\hat{H} = c(\hat{\boldsymbol{\sigma}} \cdot \hat{\mathbf{p}})$ for the Hamiltonian of the two-component theory and Eqs. (6.13b, c). For this reason, the Weyl equations, rather than the Dirac equation, were recognised as the correct equations of the neutrino only after the discovery in 1956 of the nonconservation of parity in weak interactions (where neutrinos are often involved). The revival of the Weyl theory is due to Landau, Lee and Yang and Salam[32].

By a procedure similar to the one followed in the case of the Dirac equation, it can be established that the particles described by Eqs. (10.17a, b) have spin $\frac{1}{2}$ and that the operator corresponding to spin is $\hat{\mathbf{s}} = (\hbar/2)\hat{\boldsymbol{\sigma}}$. Thus, the eigenvalues of the operator,

---

31. Neutrino is the name given by Fermi to the massless, chargeless, spin $\frac{1}{2}$ particles whose existence was postulated in 1931 by Pauli in order to account for the 'missing' energy and angular momentum in beta decay. Neutrinos were detected experimentally in 1956.

32. Landau, L.D. *Nuclear Physics*, **3**, 127 (1957);
Lee, T.D. and Yang, C.N. *Phys. Rev.* **105**, 1671 (1957);
Salam, A. *Nuovo Cimento*, **5**, 299 (1957).

$$\hat{h} = \frac{(\hat{\sigma} \cdot \hat{p})}{|\hat{p}|} \equiv \hat{\sigma}_p, \tag{10.144}$$

are proportional to the component of the spin along the direction of motion of the particle. Denoting the eigenvalue of $\hat{h}$ by h, we have, $h = \pm 1$. $\hat{h}$ is referred to as the *helicity operator* while h is the *helicity* of the particle. A particle with helicity +1 (spin parallel to $\mathbf{p}$) is called right-handed whereas one with $h = -1$ is called *left-handed.* These names arise from the resemblance of a graphical representation of the relationship between spin and momentum of the particle to a screw of the appropriate handedness (Fig. 10.4). Eqs. (10.17a) and (10.17b) thus represent spin $\frac{1}{2}$ zero-mass particles of helicity +1 and −1 respectively.[33]

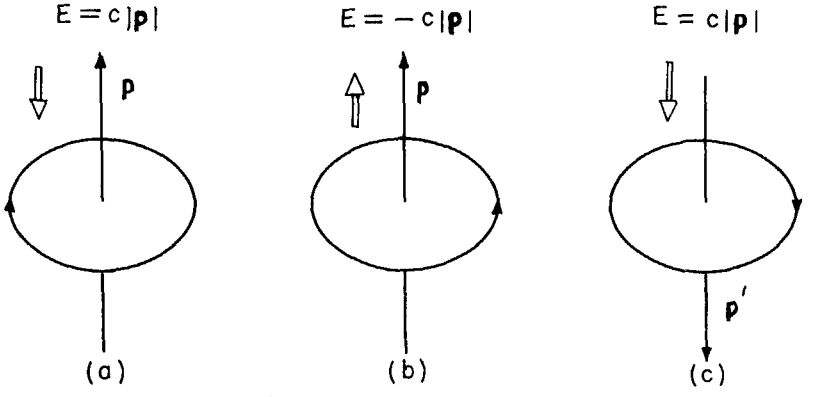


Fig. 10.4. Graphical representation of:
(a) a positive-energy (left-handed) neutrino.
(b) a negative-energy (right-handed) neutrino,
(c) an antineutrino.

Now, Eqs. (10.17a, b) were obtained on the assumption that the energy is always positive. On this basis, we will have two types of neutrinos: a right-handed one described by Eq. (10.17a) and a lefthanded one described by Eq. (10.17b), both of positive energy. However, the logic of the Dirac theory of the electron with its hole theory interpretation requires the existence of an antiparticle for every particle. In other words, negative energies should be allowed for the neutrinos. But if negative energy is permitted, both the positive and the negative square roots of Eq. (10.15) are taken care of by either of Eqs. (10.13). If we adopt (10.13[1]) as the Hamiltonian of the neutrino, then the positive energy neutrino would be right-handed and the negative energy neutrino (and, hence, the antineutrino[34]) left-handed, whereas with (10.13[2]) as the Hamiltonian we would have

33. Note that helicity is a relativistically invariant quantity only for a zero-mass particle. Since a particle with nonzero mass moves with a velocity $< c$, it is possible to find two Lorentz frames in which the helicity has opposite signs.

34. An antineutrino of momentum $\mathbf{p}'$, energy $E = c\,|\,\mathbf{p}'\,|$ and helicity +1 is the absence of a negative energy neutrino (in the negative energy sea) of energy $E = -c\,|\,\mathbf{p}\,|$, momentum $\mathbf{p} = -\mathbf{p}'$ and helicity + 1 (see Fig. 10.4).

a left-handed neutrino and a right-handed antineutrino. The experimental situation (when coupled with certain conservation laws) is that there are three (possibly more) types of neutrinos, the *electronic*, the *muonic* and the *tauonic*, but in all cases the neutrino is left-handed and the antineutrino right-handed. It follows that Eq. (10.17b), *without* the restriction to negative helicity, is the correct equation of the neutrinos. The two components of ψ correspond to the two energy states of the neutrino, but for a given sign of the energy there is only one spin state. The noninvariance of the two-component theory under space inversion is related to this absence of the other spin state. For, under space inversion a left-handed neutrino is transformed into a right handed neutrino, but no such neutrinos exist in nature. However, the theory is invariant under the combined operation of space inversion $\hat{P}$ and *charge conjugation* $\hat{C}$ (which is the replacement of particles by antiparticles) since right-handed antineutrinos exist.

Connection with Dirac Equation

It is seen from the foregoing that the states of the neutrino are the states of a zero-mass spin $\frac{1}{2}$ particle with definite helicity. Therefore, it should be possible to obtain the neutrino states from the Dirac 4-component theory corresponding to $m = 0$ by projecting out states of the appropriate helicity. From Eqs. (10.43) and (10.44), we see that the four linearly independent spinors which are the solutions of the Dirac equation with $m = 0$, are given by (where, the momentum is taken to be along the x_3-axis and where the normalising factor is neglected),

$$\begin{array}{cccc} u^{(1)} & u^{(2)} & u^{(3)} & u^{(4)} \\ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} & \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \end{array} \qquad (10.145)$$

$$\begin{array}{cccc} E = c\,|\,\mathbf{p}\,| & c\,|\,\mathbf{p}\,| & -c\,|\,\mathbf{p}\,| & -c\,|\,\mathbf{p}\,| \\ h = +1 & -1 & +1 & -1 \\ \gamma_5 = -1 & +1 & +1 & -1 \end{array} \qquad \text{(From Eq. (10.45))}$$

Below the spinors, we have listed the corresponding energy, helicity and the eigenvalue of the matrix γ_5 [defined by (10.143)]. We see that $u^{(2)}$ and $u^{(3)}$ which satisfy the relationship,

$$\gamma_5 u^{(r)} = +u^{(r)}, \qquad (10.146a)$$

have the properties required of the neutrino states, namely negative helicity when $E > 0$ and positive helicity when $E < 0$. The relationship between h and the eigenvalues of γ_5 given in (10.145) is a general one. This follows from the Dirac Hamiltonian for zero-mass particles, namely [see Eqs. (10.21[2]), (10.26) and (10.143)].

$$\hat{H}_D = c(\hat{\boldsymbol{\alpha}} \cdot \hat{\mathbf{p}}) = -c\hat{\gamma}_5(\hat{\boldsymbol{\sigma}}^D \cdot \hat{\mathbf{p}}) = -c\,|\,\mathbf{p}\,|\,\gamma_5 \hat{\sigma}_{\mathbf{p}}^D. \qquad (10.147)$$

Therefore, if $\psi(x)$ is an arbitrary solution of the zero-mass Dirac equation $\gamma_\mu \partial_\mu \psi(x) = 0$, then the function,

$$\psi_\nu(x) = \frac{1}{2}(I + \gamma_5)\psi(x), \tag{10.146b}$$

will correspond to the neutrino states. In the representation (10.30) and (10.143), ψ_ν is a 4-component wavefunction eventhough there are only two states satisfying (10.146b). ψ_ν can be reduced to an essentially two-component wavefunction by choosing a representation in which

$$\gamma_5' = \gamma_4 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad \gamma_4' = -\gamma_5 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix},$$

$$\gamma' = \gamma = \begin{pmatrix} 0 & -i\hat{\sigma} \\ i\sigma & 0 \end{pmatrix} \tag{10.148}$$

where γ is the vector whose components are γ_1, γ_2 and γ_3. That the representation (10.148) indeed corresponds to the Weyl theory can be seen from the following:

Eqs. (10.17a, b) could be combined into the single equation,

$$i\hbar\, \partial_t \Psi = \hat{H}_w \Psi, \tag{10.149}$$

with

$$\Psi = \begin{pmatrix} \Phi \\ \psi \end{pmatrix}$$

and

$$\hat{H}_w = c \begin{pmatrix} \sigma & 0 \\ 0 & -\sigma \end{pmatrix} \cdot \hat{\mathbf{p}} = i\, c\, \gamma_5 (\gamma \cdot \hat{\mathbf{p}}) \tag{10.150}$$

Comparing $\hat{H}_w$ with (10.147) which can be written as,

$$\hat{H}_D = ic\gamma_4(\gamma \cdot \hat{\mathbf{p}}), \tag{10.147'}$$

we see that, γ_5 in the Weyl theory plays the role of γ_4 in the Dirac theory. That is,

$$(\gamma_5)_{\text{Weyl}} \equiv \gamma_5' = (\gamma_4)_{\text{Dirac}}$$

Then, from

$$\gamma_1'\gamma_2'\gamma_3'\gamma_4' = \gamma_5',$$

we get,

$$\gamma_4' = -\gamma_1'\gamma_2'\gamma_3'\gamma_5' = -\gamma_1\gamma_2\gamma_3\gamma_4 = -\gamma_5.$$

10.3 THE SECOND ORDER WAVE EQUATIONS

An off-shoot of the hole theory interpretation of the Dirac equation is the suggestion that the second order Klein-Gordon Equation (10.6^2) also might be brought under relativistic quantum mechanics by a proper reinterpretation of the wavefunction. In this section, we will briefly discuss the nature of this reinter-

pretation and some of the conclusions following therefrom[35]. We will also obtain the second-order wave equation corresponding to zero-mass particles.

A. The Klein-Gordon Equation

The major reason for rejecting the Klein-Gordon equation as a quantum mechanical wave equation was the possibility of negative value for ρ defined by Eq. (10.11a). We overcome this difficulty by multiplying the right sides of Eqs. (10.11) and (10.11a) by e (where $-e$ = charge of the electron), and interpreting $\mathbf{j}$ and ρ as the *electric current* density and the *charge* density, respectively.

$$\mathbf{j} = (e\hbar/2mi)(\Phi^*\nabla\Phi - \Phi\nabla\Phi^*), \tag{10.151}$$

$$\rho = (ie\hbar/2mc^2)(\Phi^*\partial_t\Phi - \Phi\partial_t\Phi^*). \tag{10.152}$$

Eq. (10.6^2) admits plane wave solutions of the type

$$\Phi(\mathbf{r}, t) = A \exp[(i/\hbar)(\mathbf{p}\cdot\mathbf{r} - Et)], \tag{10.153}$$

with $E = \pm \epsilon_p$,

where

$$\epsilon_p = \sqrt{p^2c^2 + m^2c^4}. \tag{1.154}$$

Substituting (10.153) in (10.152), we get,

$$\rho = \frac{eE}{mc^2}|\Phi|^2, \tag{10.155}$$

or

$$\rho_+ = e(\epsilon_p/mc^2)|\Phi_+|^2;$$

$$\rho_- = -e(\epsilon_p/mc^2)|\Phi_-|^2. \tag{10.155a}$$

We interpret this result to mean that the state Φ_+ with $E = +\epsilon_p$ corresponds to particles with charge $+e$ while Φ_- with $E = -\epsilon_p$ represents particles with charge $-e$. Also, from (10.152), we see that $\rho = 0$ when Φ is real. Therefore, the Klein-Gordon equation with a real wavefunction represents neutral particles. Moreover, since $(-\kappa^2)$ is a Lorentz-invariant, Eq. (10.6^2) would be covariant under Lorentz transformations if $\Phi(x)$ is either a scalar or a pseudo scalar. Now, a scalar or a pseudo scalar wavefunction represents a spin-zero particle. Thus, (10.6^2) is the wave equation of spin-zero particles, both charged and neutral. Pions and kaons which are spin-zero mesons[36] with charges $+e, 0$ and $-e$, could be identified with

35. A more detailed discussion of the Klein-Gordon equation from the viewpoint of relativistic quantum mechanics can be found in Ref. 3, Sections 54-59. Also, see Ref. 4, Chapter 3, for a different approach to the problem. The particles described by the *K-G* equation are strong-interacting mesons. Their interaction with other particles involves creation and annihilation of particles and as such, is best treated within the framework of a quantum field theory. The discussion here is confined to the free-particle case.

36. Mesons are bosons with a mass intermediate between the electron and the nucleon. Pions (denoted by π^+ and π^0) have a mass of about 300 electron masses while kaons (K^+, K^0 and $\overline{K^0}$) are more than three times heavier than the pions.

the Klein-Gordon particles. It is found that both pions and kaons have negative intrinsic parities so that $\Phi(x)$ in (10.6[2]) is a pseudo scalar.

According to the above interpretation, charge $+e$ corresponds to the particle and charge $-e$ to the antiparticle. Unlike in the case of the Dirac theory, the antiparticle need not be viewed as a hole in a negative energy sea but could be regarded as a positive energy particle with charge $-e$ [see Eqs. (10.152[1]), (10.162) and (10.165b) below]. An important consequence of this is that production (or destruction) of the pions need not be in particle-antiparticle pairs.

The foregoing treatment is somewhat oversimplified. For, solutions corresponding to either positive energy *or* negative energy alone do not form a Hilbert space (with the normal definition of the scalar product) except in the nonrelativistic limit. Therefore, a general solution of the Klein-Gordon equation would be a linear combination of Φ_+ and Φ_-. Alternatively, we can write the wavefunction as

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \tag{10.156}$$

where , $\psi_1 \to 0$, in the non-relativistic limit, for negative energy while $\psi_2 \to 0$ for positive energy. We have[37],

$$\psi_1 = \frac{1}{2}\left(\Phi + \frac{i\hbar}{mc^2}\partial_t\Phi\right) = \frac{1}{2}\left(1 + \frac{E}{mc^2}\right)\Phi, \tag{10.157a}$$

$$\psi_2 = \frac{1}{2}\left(\Phi - \frac{i\hbar}{mc^2}\partial_t\Phi\right) = \frac{1}{2}\left(1 - \frac{E}{mc^2}\right)\Phi, \tag{10.157b}$$

Then,

$$\psi_1 + \psi_2 = \Phi;\ mc^2(\psi_1 - \psi_2) = i\hbar\partial_t\Phi. \tag{10.158}$$

From (10.157a) and (10.158), we get,

$$i\hbar\partial_t\psi_1 = \frac{1}{2}\left[i\hbar\partial_t\Phi - \frac{\hbar^2}{mc^2}\partial_t^2\Phi\right]$$

$$= \frac{1}{2}\left[mc^2(\psi_1 - \psi_2) - \frac{\hbar^2}{mc^2}\left(c^2\nabla^3 - \frac{m^2c^4}{\hbar^2}\right)(\psi_1 + \psi_2)\right], \text{ by Eq. (10.6}^2\text{).}$$

That is,

$$\left.\begin{aligned} i\hbar\partial_t\psi_1 &= -\frac{\hbar^2}{2m}\nabla^2(\psi_1 + \psi_2) + mc^2\psi_1 \\ \text{Similarly,} \qquad & \\ i\hbar\partial_t\psi_2 &= \frac{\hbar^2}{2m}\nabla^2(\psi_1 + \psi_2) - mc^2\psi_2 \end{aligned}\right\} \tag{10.159[1]}$$

Introducing the matrices,

37. Note that, by virtue of the second order character of the Klein-Gordon equation, Φ and $\partial_t\Phi$ are linearly independent.

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{10.160}$$

Eq. (10.159^1) can be written as,

$$i\hbar\partial_t \Psi = \hat{H}_{KG}\Psi, \tag{10.159^2}$$

with

$$\hat{H}_{KG} = (\hat{\tau}_3 + i\hat{\tau}_2)\frac{\hat{\mathbf{p}}^2}{2m} + mc^2\hat{\tau}_3. \tag{10.161}$$

(10.159^2) is the Hamiltonian form of the Klein-Gordon equation. Multiplying this equation by $(i\hbar\partial_t + \hat{H}_{KG})$ its equivalence to (10.6^2) is easily established.

The matrices (10.160) are seen to be identical with the Pauli spin matrices (5.34) and, therefore, obey the same algebra as the latter. However, the space in which these matrices are defined is the *charge* or *isospin space* (and not the angular momentum space of the σ_k's). Therefore, the two discrete degrees of freedom[38] implied by the components of Ψ are the two charge states of a Klein-Gordon particle with a complex wavefunction.

In terms of Ψ, we have from (10.152), (10.158), (10.156), and (10.160),

$$\rho = e(\psi_1^*\psi_1 - \psi_2^*\psi_2) = e\Psi^\dagger\tau_3\Psi. \tag{10.152^1}$$

The equation of continuity (10.9^2) ensures that the total charge $\int \rho d^3\mathbf{r}$ is a constant. In the case of a single particle, the total charge should be $\pm e$. Correspondingly, we have,

$$\int \Psi^\dagger\tau_3\Psi d^3\mathbf{r} \equiv (\Psi, \hat{\tau}_3\Psi) = \pm 1. \tag{10.162}$$

Eq. (10.162) amounts to a modification of the standard definition [Eq. (2.33)] of the scalar product: the scalar product of two vectors Ψ and Φ is $(\Psi, \hat{\tau}_3\Phi)$, so that the norm of Ψ is given by

$$||\Psi||^2 = (\Psi, \hat{\tau}_3\Psi). \tag{10.163a}$$

Similarly, the expectation value of an operator $\hat{A}$ is given by

$$\langle\hat{A}\rangle = (\Psi, \hat{\tau}_3\hat{A}\Psi). \tag{10.163b}$$

Whereas the standard definition leads to positive normalization of vectors and real expectation values for Hermitian operators, the modified definition permits negative normalization for vectors and imaginary expectation values for Hermitian operators. In fact, the expectation value of a Hermitian operator would be real or imaginary depending on whether the operator commutes or anticommutes with $\hat{\tau}_3$. It is, however, possible (as well as desirable) to modify the definition of the Hermitian operator also so that its expectation value as defined by (10.163b) is always real. Thus, $\hat{A}$ is Hermitian if

---

38. Whenever a system has degrees of freedom other than those connected with spatial co-ordinates, its wavefunction can be written as a column matrix with two or more elements. Conversely, a wavefunction in the form of a column matrix implies the existence of degrees of freedom connected with discrete variables.

$$(\Psi, \hat{\tau}_3 \hat{A} \Psi)^* = (\hat{\tau}_3 \hat{A} \Psi, \Psi) = (\Psi, \hat{A}^\dagger \hat{\tau}_3 \Psi) = (\Psi, \hat{\tau}_3 \hat{A} \Psi),$$

That is, if

$$\hat{A} \equiv \hat{\tau}_3 \hat{A}^\dagger \hat{\tau}_3 = \hat{A}. \tag{10.164}$$

We note that the Hamiltonian (10.161) is Hermitian according to this new definition, but not Hermitian according to the standard definition (For an operator that commutes with $\hat{\tau}_3$, the new and the old definitions are the same).

The above results indicate that certain concepts of nonrelativistic quantum mechanics have to be modified before they can make sense in relativistic quantum mechanics. For example, one of the basic concepts of non-relativistic quantum mechanics is that the results of possible measurements on a system are the eigenvalues of certain Hermitian operators (Postulate 1, Chapter 3). In particular, the eigenvalues of the Hamiltonian represent the measurable energies of the system. Diagonalising the matrix,

$$H_{KG} = (\tau_3 + i\tau_2)\frac{\mathbf{p}^2}{2m} + mc^2\tau_3$$

$$= \begin{pmatrix} \left(\frac{\mathbf{p}^2}{2m} + mc^2\right) & \frac{\mathbf{p}^2}{2m} \\ \frac{-\mathbf{p}^2}{2m} & -\left(\frac{\mathbf{p}^2}{2m} + mc^2\right) \end{pmatrix}$$

we see that the eigvenvalues of the Hamiltonian $\hat{H}_{KG}$ are

$$E_\lambda = \lambda \in_p, \quad \lambda = + \text{ or } -. \tag{10.165a}$$

But the negative eigenvalue E_- cannot represent a measurable energy of the Klein-Gordon particle. On the other hand, according to Eq. (10.163b) the measurable energies are given by

$$E = (\Psi_\lambda, \hat{\tau}_3 \hat{H}_{KG} \Psi_\lambda) = E_\lambda (\Psi_\lambda, \hat{\tau}_3 \Psi_\lambda) = \lambda E_\lambda = \in_p > 0. \tag{10.165b}$$

Here, the equation $\hat{H}_{KG}\Psi_\lambda = E_\lambda \Psi_\lambda$ and the fact (see Problem 10.14 below) that $(\Psi_\lambda, \hat{\tau}_3 \Psi_\lambda) = \lambda$, have been used. The modified scalar product is responsible for the sensible result (10.165b). The positive definiteness of $< \hat{H}_{KG} >$ also follows quite generally from the expression,

$$\int \Psi^\dagger \hat{\tau}_3 \hat{H}_{KG} \Psi d^3\mathbf{r} = \int \left\{ (\psi_1^* + \psi_2^*) \frac{\mathbf{p}^2}{2m} (\psi_1 + \psi_2) + mc^2 (\psi_1^* \psi_1 + \psi_2^* \psi_2) \right\} d^3\mathbf{r}. \tag{10.165c}$$

It is interesting to note that whereas in the Klein-Gordon theory the energy is positive definite and the norm (of vectors) is not positive definite, the reverse conditions obtain in the Dirac theory.

There is a certain similarity between (10.161) and the Dirac Hamiltonian (10.21[2]) with $\hat{\tau}_3$ taking the place of $\hat{\beta}$. It follows that it should be possible to find a representation, analogous to the Foldy-Wouthuysen representation of the Dirac theory, in which the positive and the negative *charge states* are separated. In this new representation, known as the *Feshbach-Villars representation*[39], the Hamiltonian would be given by [cf. Eq. (10.87[1])],

39. Feshbach, H. and Villars, F. *Revs. Mod. Phys.* **30**, 24 (1958).

$$\hat{H}'_{KG} = \hat{U}\hat{H}_{KG}\hat{U}^{-1} = \hat{\tau}_3 \in_p . \tag{10.166}$$

By a procedure similar to that adopted in the case of the Dirac equation, it can be shown that

$$\hat{U} = \frac{(\hat{1} + \hat{\tau}_1) + (\in_p + mc^2)}{2(mc^2 \in_p)^{1/2}} \tag{10.167}$$

$\hat{U}$ satisfies the relationship, $\hat{u} \equiv \hat{\tau}_3 \hat{U}^+ \hat{\tau}_3 = \hat{U}^{-1}$, and is hence unitary according to the modified definition of a unitary operator.

Problem 10.14: Assuming Ψ to be of the form,

$$\Psi = \frac{1}{\sqrt{V}} \begin{pmatrix} \psi_{10} \\ \psi_{20} \end{pmatrix} \exp\left[(i/\hbar)(\mathbf{p}\cdot\mathbf{r} - Et)\right],$$

show that the normalizations +1 and −1 in (10.162) correspond to $E = +\in_p$ and $E = -\in_p$ respectively (Here, V is the volume in which the system is supposed to be enclosed).

Problem 10.15: Deduce (10.167).

B. Wave Equation of the Photon

Just as in the case of the Dirac equation, setting the mass $m = 0$ in the Klein-Gordon equation (10.6²) does *not* lead to the wave equation of the zero-mass bosons. Certain further constraints have to be imposed. In fact, the zero-mass equation is given by

$$A_\mu(x) = 0, \quad (\mu = 1, 2, 3, 4), \tag{10.168}$$

with the constraint,

$$\partial_\mu A_\mu(x) = 0, \tag{10.169}$$

where A_μ are the components of a 4-vector. Eq. (10.168) is, however, *not* a quantum mechanical wave equation but is (when combined with (10.169)) the *classical* wave equation of the free electromagnetic field. That is, Eqs. (10.168) and (10.169) are equivalent to Maxwell's equations in free space. This equivalence is seen as follows:

Maxwell's equations, in terms of the electric field **E** and the magnetic induction **B** are given by

$$\text{div } \mathbf{B} = 0, \tag{10.170a}$$

$$\text{curl } \mathbf{E} + (1/c)\partial_t \mathbf{B} = 0, \tag{10.170b}$$

$$\text{curl } \mathbf{B} - (1/c)\partial_t \mathbf{E} = (4\pi/c)\mathbf{j}, \tag{10.170c}$$

$$\text{div } \mathbf{E} = 4\pi\rho, \tag{10.170d}$$

The vector potential **A** and the scalar potential V are defined, in view of Eqs. (10.170a, b), by

$$\mathbf{B} = \text{curl } \mathbf{A},\ \mathbf{E} = -\text{grad } V - (1/c)\partial_t \mathbf{A}. \tag{10.171}$$

Substituting (10.171) in (10.170c, d) and making use of the identity, curlcarl ≡ grad div $-\nabla^2$, we get,

$$-A_\mu - \partial_\mu(\partial_\nu A_\nu) = -\frac{4\pi}{c} j_\mu, \tag{10.172}$$

where, $$A_4 = iV;\quad j_4 = ic\rho. \tag{10.173}$$

Eq. (10.168) results from imposing the subsidiary condition (10.169) and by setting $j_\mu = 0$ (free space).

In terms of the A_μ, the components of the fields (10.171) are given by

$$F_{ij} = \in_{ijk} B_k;\ F_{k4} = -iE_k, \tag{10.174}$$

where, $$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{10.175}$$

and is known as the *electromagnetic field strength tensor.*

Eq. (10.169), referred to as the *Lorentz condition*, amounts to a particular choice of the four-potential. And such a choice is permitted by the fact that only the fields, and not the potentials, are observables so that any two potentials related by the transformation,

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \Lambda, \tag{10.176}$$

(where Λ is a scalar function) which leaves the fields (10.175) invariant, are physically equivalent. Eq. (10.176) defines a *gauge transformation* while a potential satisfying Eq. (10.169) is a *Lorentz gauge*. Different Lorentz gauges can be obtained by choosing different Λ's that satisfy $\Lambda = 0$.

The constraint (10.169) as well as the imaginary nature of A_4 (since V is real) introduce certain problems for the quantum theory of the electromagnetic field. In order to avoid these difficulties the *radiation* (or *Coulomb*) gauge defined by

$$\text{div } \mathbf{A} = 0,\ A_4 = 0, \tag{10.169a}$$

has been used. But in this gauge, Eq. (10.168) becomes,

$$\mathbf{A} = 0, \tag{10.168a}$$

which is not Lorentz covariant as **A** is a 3-vector. The problems associated with the Lorentz gauge have, however, been successfully solved by Gupta[40] and Bleuler[41]. So we will base our discussion of the electromagnetic field on Eqs. (10.168) and (10.169).

Eventhough (10.168) is a classical field equation, its similarity to the quantum mechanical Klein-Gordon equation (10.6²) suggests that it could be viewed as the wave equation (with $A_\mu(x)$ as the wavefunctions) of certain zero-mass particles (the *photons*) that we can associate with the electromagnetic field. Conversely (and more fruitfully), the Klein-Gordon equation, and therefore any quantum mechanical wave equation, could be regarded as a classical field equation. The

40. Gupta, S.N. *Proc. Phys. Soc.* **A63**, 681 (1950).
41. Bleuler, K. *Helv. Phys. Acta*, **23**, 567 (1950).

technique of field quantization developed for quantizing the electromagnetic field could then be applied to these quantum mechanical wave equations. We will see in the next Chapter that such a procedure gets rid of many of the difficulties and inconsistencies that are encountered in treating the relativistic wave equations as the quantized equations of motion of point particles.

A quantum mechanical treatment of Eq. (10.168) is not particularly useful. Therefore, we will postpone to the next chapter a discussion of the equation from the view point of quantum theory.

Besides the wave equations discussed in this Chapter, there are also other relativistic wave equations both of the first and the second orders[42]. These correspond to massive (that is, non-zero mass) particles of spin ≥ 1. A discussion of these is beyond the scope of this book.

10.4 CHARGE CONJUGATION

Charge conjugation is a symmetry operation in which particles are transformed into antiparticles and *vice-versa*. Obviously, then, the electric charge changes sign under charge conjugation. But also other properties like magnetic moment and helicity (in the case of neutrinos) also will change sign under this transformation. Therefore, electrical neutriality alone will not ensure that a particle is its own antiparticle. If a particle is its own antiparticle it is called truly neutral. Neutrinos are examples of particles which are not truly neutral though electrically neutral whereas neutral pions and photons are examples of truly neutral particles. The wave function of a truly neutral particle must be an eigenstate of the *Charge Conjugation operator* $\hat{C}$. That is, if ψ is the wave function of such a particle, then

$$\psi_c = \hat{C}\psi = c\psi = \pm\psi. \tag{10.177}$$

The last part of this equation follows from the fact that a two-fold application of $\hat{C}$ should be equivalent to the identity operation, so that $c^2 = 1$. c is known as the *charge parity* of the particle. Thus neutral particles with $c = +1$ have *positive* charge parity whereas those with $c = -1$ have *negative* charge parity.

The charge parity of a particle can be determined by observing its interaction with other particles of known charge parity. For example, the charge parity of photons is known to be negative. The observation that a neutral pion decays into two photons, then, suggests that the charge parity of the pion is positive[43].

In this section, we will determine the charge conjugation operator $\hat{C}$ corresponding to each of the particles (wave equations) discussed in this chapter.

The Dirac Equation

We will start with the covariant form, Eq. (10.25[2]), of the Dirac equation. In terms of the momentum $\hat{p}_\mu = -i\hbar\partial_\mu$, the equation (in coordinate representation) reads,

---

42. See, Lurie, D. *Particles and Fields* (Interscience, 1968) Chapter 1.
43. Charge Parity, like (space-inversion) parity, is a multiplicative quantum number.

$$(\gamma_\mu p_\mu - \text{imc})\,\psi(x) = 0. \tag{10.178}$$

Since the electric charge is explicitly involved in charge conjugation, we have to consider the equation in the presence of an electromagnetic field. The latter can be represented by the 4-potential A_μ (see Section 10.3B). The effect of the field on the Dirac particle is then to change its momentum from p_μ to $[p_\mu - (e/c)A_\mu]$. Thus, the Dirac particle in an electromagnetic field is represented by the equation,

$$[\gamma_\mu\{p_\mu - (e/c)A_\mu\} - \text{imc}]\psi(x) = 0. \tag{10.179}$$

The *charge conjugation state* ψ_c must, then, satisfy the equation,

$$[\gamma_\mu\{p_\mu + (e/c)A_\mu\} - \text{imc}]\psi_c(x) = 0. \tag{10.180}$$

We have to find the operator that transforms ψ to ψ_c:

Taking the complex conjugate of Eq. (10.179), we have,

$$[\gamma_\mu^*\{p_\mu^* - (e/c)A_\mu^*\} + \text{imc}]\psi^*(x) = 0. \tag{10.181^1}$$

But

$$p_k^* = -p_k,\ (k = 1,2,3);\ p_4^* = p_4, \tag{10.182}$$

$$A_k^* = p_k,\ (k = 1,2,3);\ A_4^* = -A_4, \tag{10.183}$$

Substituting from (10.182) and (10.183) in (10.181^1) and defining,

$$\psi^* = \hat{B}\psi_c, \tag{10.184^1}$$

we get,

$$[\hat{\gamma}^* \cdot \{\hat{\mathbf{p}} + (e/c)\mathbf{A}\} - \hat{\gamma}_4^*\{\hat{p}_4 + (e/c)A_4\} - \text{imc}]\hat{B}\psi_c = 0, \tag{10.181^2}$$

where $\hat{\gamma}$ is defined by Eq. (10.148).

Multiplying Eq. (10.181^2) from the left by $\hat{B}^{-1}$, we have[44],

$$[\hat{\gamma}^{*\prime} \cdot \{\hat{\mathbf{p}} + (e/c)\mathbf{A}\} - \hat{\gamma}_4^*\{\hat{p}_4 + (e/c)A_4\} - \text{imc}]\psi_c = 0, \tag{10.185}$$

where

$$\hat{\gamma}_\mu^{*\prime} = \hat{B}^{-1}\hat{\gamma}_\mu^*\hat{B}. \tag{10.186}$$

Eq. (10.185) will agree with Eq. (10.180) if

$$\hat{B}^{-1}\hat{\gamma}_k^*\hat{B} = \hat{\gamma}_k, \text{ and } \hat{B}^{-1}\hat{\gamma}_4^*\hat{B} = -\hat{\gamma}_4. \tag{10.187^1}$$

But if we use the representation (10.30a, c) for the $\hat{\gamma}_\mu$'s, then,

$$\hat{\gamma}_\mu^* = -\hat{\gamma}_\mu, (\mu = 1,3);\ \hat{\gamma}_\mu^* = \hat{\gamma}_\mu, (\mu = 2,4),$$

so that Eq. (10.187^1) implies,

$$\hat{B}\hat{\gamma}_\mu = -\hat{\gamma}_\mu\hat{B},\ (\mu = 1,3,4);\ \hat{B}\hat{\gamma}_2 = \hat{\gamma}_2\hat{B}. \tag{10.187^2}$$

That is, $\hat{B}$ is an operator that commutes with $\hat{\gamma}_2$ but anticommutes with the other $\hat{\gamma}_\mu$'s. It is obvious that $\hat{B}$ should be proportional to $\hat{\gamma}_2$, and there is no loss of generality in identifying $\hat{B}$ with $\hat{\gamma}_2$. Then from (10.184^1), we have,

$$\psi_c = \hat{\gamma}_2\psi^* = \hat{\gamma}_2\hat{K}_c\psi; \tag{10.184^2}$$

---

44. Note that $\hat{B}$, being independent of the space-time co-ordinates, should commute with $\hat{\mathbf{p}}$ and $\mathbf{A}$.

where K_c is the complex conjugation operator. Thus, the charge conjugation operator $\hat{C}$ is given by

$$\hat{C} = \hat{\gamma}_2 \hat{K}_c. \tag{10.188}$$

$\hat{C}$, like the time-reversal operator $\mathcal{T}$(Eq. 10.138[2]), is an *antiunitary* operator. From Eqs. (10.187[2]) and (10.188), we have,

$$\hat{C}^{-1}\hat{\gamma}_\mu \hat{C} = \begin{cases} \hat{\gamma}_\mu, & (\mu = 1, 2, 5), \\ -\gamma_\mu, & (\mu = 3, 4), \end{cases} \tag{10.187[3]}$$

where $\hat{\gamma}_5$ is defined by Eq. (10.143).

If ψ represents electrons, ψ_c will represent positrons. If we write $\psi = \begin{pmatrix} \phi \\ \chi \end{pmatrix}$, then, from (10.184[2]) and (10.30c), we have,

$$\psi_c = \begin{pmatrix} -i\sigma_2 \chi^* \\ i\sigma_2 \phi^* \end{pmatrix} \tag{10.189}$$

In the non-relativistic case, we have [see Eqs. (10.67) and (10.68)],

$$\psi \to \phi; \ \psi_c \to i\sigma_2 \psi^* = \mathcal{T}\phi, \tag{10.189a}$$

where $\mathcal{T}$ is the time-reversal operator [Eq. (10.138[(1)])]. That is, *the positron could be viewed as an electron going backward in time*[45] (see Fig. 11.3). Such a viewpoint will be valid even in the relativistic case where the positive and negative energy states are separated [Foldy-Wouthuysen representation : see Eqs. (10.95a, b)].

Klein-Gordon Equation

The electric charge of the Klein-Gordon particle is positive or negative according as the integral $\int \psi^\dagger \tau_3 \psi d^3\mathbf{r}$ is positive or negative [Eq. (10.162)]. Writing

$$\psi = (1/V)^{\frac{1}{2}} \begin{pmatrix} \psi_{10} \\ \psi_{20} \end{pmatrix} \exp\left[(i/\hbar)(\mathbf{p}\cdot\mathbf{r} - Et)\right], \tag{10.190}$$

the condition is

$$\psi_{10}^* \psi_{10} - \psi_{20}^* \psi_{20} = \pm 1. \tag{10.191[1]}$$

Here V is the volume in which the system is enclosed. But, according to Eq. (10.155a), the sign of the electric charge depends on the sign of the energy. Therefore, which of the two conditions in Eq. (10.191[1]) is realized should depend on whether the E in Eq. (10.190) is positive or negative. In fact, we will show below that the +1 (positive charge) in Eq. (10.191[1]) corresponds to $E > 0$ and the −1 (negative charge) to $E < 0$.

From Eqs. (10.156) and (10.190), we have,

---

45. The concept of the positron as an electron going backward in time was introduced by Feynman (Footnote 16, Chapter 11).

$$\psi_1 = (1/V)^{\frac{1}{2}}\psi_{10} \exp\left[(i/\hbar)(\mathbf{p}\cdot\mathbf{r} - Et)\right],$$

$$\psi_2 = (1/V)^{\frac{1}{2}}\psi_{20} \exp\left[(i/\hbar)(\mathbf{p}\cdot\mathbf{r} - Et)\right]. \tag{10.192}$$

Substituting from (10.192) in (10.159[1]), we get,

$$(E - mc^2)\psi_{10} = (\mathbf{p}^2/2\,\mathrm{m})(\psi_{10} + \psi_{20}), \tag{10.193a}$$

$$(E + mc^2)\psi_{20} = -(\mathbf{p}^2/2\,\mathrm{m})(\psi_{10} + \psi_{20}), \tag{10.193b}$$

or

$$\frac{\psi_{10}}{\psi_{20}} = \frac{mc^2 + E}{mc^2 - E}. \tag{10.194}$$

Case 1: $E = \varepsilon_p > 0$.

$$\left|\frac{\psi_{10}}{\psi_{20}}\right| = \frac{mc^2 + \varepsilon_p}{|mc^2 - \varepsilon_p|} > 1, \tag{10.194a}$$

which means that in Eq. (10.191[1]) the upper sign has to be chosen. Then from Eq. (10.194a) we obtain,

$$\psi_{10} = \frac{mc^2 + \varepsilon_p}{2\sqrt{mc^2\varepsilon_p}} = \phi_0, \text{ say;} \tag{10.195a}$$

$$\psi_{20} = \frac{|mc^2 - \varepsilon_p|}{2\sqrt{mc^2\varepsilon_p}} = \chi_0. \tag{10.195b}$$

Case 2: $E = -\varepsilon_p < 0$.

In this case

$$\frac{|\psi_{10}|}{|\psi_{20}|} = \frac{|mc^2 - \varepsilon_p|}{mc^2 + \varepsilon_p} < 1, \tag{10.194b}$$

so that we have to choose the lower sign in eq. (10.191[1]). Then, in place of (10.195a, b) we get,

$$\psi_{10} = \chi_0 \;;\; \psi_{20} = \phi_0. \tag{10.196}$$

Thus, we see from Eqs. (10.190), (10.191[1]), (10.195a, b) and (10.196), that if [cf. Eq. (10.156)]

$$\psi = \begin{pmatrix}\phi\\ \chi\end{pmatrix} \tag{10.190a}$$

represents a positive charge state of the Klein-Gordon particle, then

$$\psi_c = \begin{pmatrix}\chi^*\\ \phi^*\end{pmatrix} \tag{10.190b}$$

represents the negative charge state of the particle, where,

$$\phi = (1/V)^{\frac{1}{2}} \phi_0 \exp[(i/\hbar)(\mathbf{p}\cdot\mathbf{r}) - \epsilon_p t)], \tag{10.197a}$$

$$\chi = (1/V)^{\frac{1}{2}} \chi_0 \exp[(i/\hbar)(\mathbf{p}\cdot\mathbf{r}) - \epsilon_p t)], \tag{10.197b}$$

and

$$\phi_0^*\phi_0 - \chi_0^*\chi_0 \equiv \psi_0^\dagger \tau_3 \psi_0 = +1, \tag{10.191^2}$$

with

$$\psi_0 = \begin{pmatrix} \phi_0 \\ \chi_0 \end{pmatrix}. \tag{10.198}$$

Hence the charge conjugation operator $\hat{C}$, defined by $\psi_c = \hat{C}\psi$, is given by

$$\hat{C} = \hat{\tau}_1 \hat{K}_c. \tag{10.199}$$

[Note that $\hat{K}_c$ will change $\hat{\mathbf{p}}$ to $-\hat{\mathbf{p}}$ {see Eq. (B.19a)}].

It is seen from (10.190a, b) that

$$(\psi_c, \hat{\tau}_3 \psi) = 0, \tag{10.200}$$

so that ψ and ψ_c are orthogonal to each other (in the τ-space) according to the modified definition (10.163a) of the scalar product. But, for a neutral Klein-Gordon particle, we have,

$$||\psi||^2 = (\psi, \hat{\tau}_3 \psi) = 0. \tag{10.201}$$

That is, the neutral particle is represented by a *null vector* in the τ-space. Hence the state of a neutral Klein-Gordon particle is invariant under charge conjugation, which implies that a neutral pion is its own antiparticle (truly neutral).

Problem 10.15: If Φ_+, Φ_- and Φ_0 are the wave functions of the Klein-Gordon particle [Eq. (10.155)] with charge $+e, -e$ and 0 respectively, show that $\Phi_- = \Phi_+^*$ and $\Phi_0 = \Phi_0^*$.

The Zero-Mass Particles

Eqs (10.170c, d) and (10.171) show that A_μ changes sign under charge conjugation (which changes the sign of the electric charge and current):

$$\hat{C} A_\mu = -A_\mu \tag{10.202}$$

Thus, A_μ represents truly neutral particles (photons) with negative charge parity.

Using the Dirac representation (10.146b) for the neutrino wave function, we see that the operation of charge conjugation alone [represented by (10.188)] does not lead to a state of the antineutrino. However, a combined operation of space inversion [Parity operation, represented by (10.137)] and charge conjugation will transform the state of a neutrino into a state of the antineutrino:

$$\hat{C}\hat{P}\psi_\nu = \psi_{\bar{\nu}}. \tag{10.203}$$

REFERENCES

1. Björken J.D. and Drell, S.D. *Relativistic Quantum Mechanics* (McGraw-Hill, New York 1964), Chapters 1 to 5.
2. Schiff, L.I. *Quantum Mechanics*, 3rd ed. (McGraw-Hill, 1968), Chapter 13.
3. Davydov, A.S. *Quantum Mechanics* (Pergamon Press, 1985) Translation by D. ter Haar, Chapter VIII.
4. Schweber, S.S. *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York 1961), Chapters 3 to 5.

CHAPTER 11

ELEMENTS OF FIELD QUANTIZATION

11.1 INTRODUCTION

We have remarked in the last chapter that the relativistic wave equations could be viewed as classical field equations and that the quantization of these field equations might lead to a correct relativistic quantum theory of the underlying particles. In this Chapter, we propose to discuss some of the concepts and methods underlying the quantization of fields. The treatment would be somewhat sketchy. In fact, our attempt in this regard could be likened to that of a traveller who having traversed a continent, crossing several rivers and mountains in the process, has finally come to the shore of a vast and deep ocean. Being ill-equipped as well as too tired to continue the travel into the sea, he contents himself with having a view of the ocean from the shore. It is hoped that an account of this 'view from the shore' will inspire the more adventurous among the readers to undertake an exploration into this vast ocean that is quantum field theory.

11.2 LAGRANGIAN FIELD THEORY

What we are going to describe in this chapter comes under *Lagrangian* (or *Canonical*) field theory as distinguished from *axiomatic* field theory[1]. As the name implies, the Lagrangian field theory is based on the Lagrangian-Hamiltonian canonical formulation of classical mechanics. The main steps in this formulation are the following[2].

*CM*1. Choose a set of generalized coordinates $\{q_i\}$ for the system.

*CM*2. Set up a Lagrangian function[3],

$$L \equiv L(q, \dot{q}, t), \tag{11.1}$$

---

1. For an account of the axiomatic field theory, see Roman, P. *Introduction to Quantum Field Theory* (John Wiley, New York 1969), Part II.
2. See, for example, Landau and Lifshitz, *Mechanics* (Pergamon Press 1969), Chapters I and VII.
3. For a free system, L would be independent of t.

where $\dot{q}_i = \dfrac{\partial q_i}{\partial t}$, are the generalized velocities and t denotes the time. Also, q stands for the set $\{q_i\}$.

*CM*3. The action integral S_{21} between times t_1 and t_2 is defined by

$$S_{21} = \int_{t_1}^{t_2} L(q, \dot{q}, t)dt. \tag{11.2}$$

The condition that the change $\delta S_{21} = 0$ corresponding to a variation δq_i in q_i subject to the constraint $\delta q_i(t_1) = \delta q_i(t_2) = 0$, then leads to the Euler-Lagrange equations,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0. \tag{11.3}$$

*CM*4. Define a Hamiltonian function H by

$$H \equiv H(q, p, t) = \Sigma_i p_i \dot{q}_i - L, \tag{11.4}$$

where p_i is the momentum conjugate to the co-ordinate q_i, and is given by

$$p_i = \frac{\partial L}{\partial \dot{q}_i}. \tag{11.5}$$

Then from Eq. (11.3), it follows that

$$\dot{p}_i = \frac{\partial L}{\partial q_i}. \tag{11.6}$$

Substitute for L in (11.2) from (11.4). Again setting $\delta S_{21} = 0$ corresponding to independent variations δq_i and δp_i in q_i respectively, one obtains Hamilton's canonical equations,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}. \tag{11.7}$$

From this, the equation of motion for a general dynamical variable $F \equiv F(q, p, t)$ is easily deduced:

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + [F, H]_{PB}, \tag{11.8}$$

where,

$$[F, H]_{PB} = \Sigma_i \left(\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial q_i}\frac{\partial F}{\partial p_i}\right), \tag{11.9}$$

and is known as the *Poisson bracket* of F and H.

Canonical Quantization

In the foregoing, the number of generalized co-ordinates is equal to the number of independent degrees of freedom of the system. For a mechanical system of particles, this number is finite. The quantization of such a system is done by the following procedure (described already in Section 3.1):

*CQ*1. Replace all dynamical variables by corresponding Hermitian operators in accordance with Postulate IV of Section 3.1.

*CQ*2. Replace the Poisson bracket in (11.8) by the commutator bracket $[\hat{F}, \hat{H}] \equiv \hat{F}\hat{H} - \hat{H}\hat{F}$ upon $i\hbar$.

*CQ*3. Prescribe algebraic relations for the basic canonical operators $\hat{q}_i$ and $\hat{p}_i$.

For a mechanical system, the algebraic relations for the co-ordinates and momenta are the Heisenberg commutation rules, (3.12a, b) that follow[4] directly from prescription CQ2. However, these relations could be different for a field.

Coordinates of the Field

Now, the basic difference between a mechanical system and a field is that the latter has infinite number of degrees of freedom. A field is specified by its amplitudes at all points of space. Moreover, the amplitudes at different space points are independent of each other. Thus, the amplitudes $\psi(\mathbf{r}, t)$ play the same role in the case of a field as the generalized co-ordinates $q_i(t)$ in the case of a mechanical system. Obviously then, since there are infinite number of space points, the number of degrees of freedom represented by $\psi(\mathbf{r}, t)$ is infinite.

The Classical Field Equations

Except for some modifications entailed by the fact that the coordinates of the field are themselves functions of the space coordinates, the procedure outlined above for the case of mechanical systems can be adopted both for obtaining the classical field equations and for their quantization. The dependence of ψ on $\mathbf{r}$ which is a continuous variable, necessitates two types of modifications. One is the necessity to introduce a *Lagrangian density* $\mathcal{L}$. The Lagrangian L of the field would be an integral of $\mathcal{L}$ over space. The other is that the Lagrangian density would have to be a function not only of ψ, $\dot{\psi}$ and t but also of grad ψ. Thus,

$$\mathcal{L} \equiv \mathcal{L}(\psi, \text{ grad } \psi, \dot{\psi}, t), \tag{11.10}$$

$$L = \int_V \mathcal{L} d^3\mathbf{r}, \tag{11.11}$$

and

$$S_{21} = \int_{t_1}^{t_2} L \, dt = \int_{t_1}^{t_2} dt \int^V \mathcal{L}(\psi, \text{ grad } \psi, \dot{\psi}, t) d^3\mathbf{r}. \tag{11.12}$$

Here V denotes the normalization volume. The change δS_{21} corresponding to an infinitesimal variation $\delta\psi$ in ψ, with $\delta\psi(\mathbf{r}, t_1) = \delta\psi(\mathbf{r}, t_2) = 0$, is given by

$$\delta S_{21} = \int_{t_1}^{t_2} dt \int_V \delta\mathcal{L} d^3\mathbf{r} = \int_{t_2}^{t_1} \delta L \, dt. \tag{11.13[1]}$$

---

4. From (11.9), the Poisson brackets for q_i and p_i are deduced to be,

$$[q_i, p_j]_{PB} = \delta_{ij}, [q_i, q_j]_{PB} = [p_i, p_j]_{PB} = 0.$$

$$\text{Now, } \delta\mathcal{L}(\psi, \text{grad}\,\psi, \dot{\psi}, t) = \frac{\partial \mathcal{L}}{\partial \psi}\delta\psi + \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)} \cdot \delta(\text{grad}\,\psi) + \frac{\partial \mathcal{L}}{\partial \dot{\psi}}\delta\dot{\psi}$$

(11.14)

$$= \frac{\partial \mathcal{L}}{\partial \psi}\delta\psi + \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)} \cdot \text{grad}\,(\delta\psi) + \frac{\partial \mathcal{L}}{\partial \dot{\psi}}\frac{\partial}{\partial t}(\delta\psi).$$

Substituting (11.14) in (11.13[1]) and making use of the results,

$$\int_V \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)} \cdot \text{grad}\,(\delta\psi) d^3\mathbf{r} = \sum_{x,y,z} \int\int \left\{ \int \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)_x} \frac{\partial}{\partial x}(\delta\psi) \cdot dx \right\} dy\,dz$$

$$= \sum_{x,y,z} \left\{ \int\int \left(\frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)_x} \cdot \delta\psi \right) dy dz - \int_V \frac{\partial}{\partial x} \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)_x} \delta\psi d^3\mathbf{r} \right\}$$

$$= -\int_V \text{div} \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)}(\delta\psi) d^3\mathbf{r}, \tag{11.15}$$

(where the surface integral vanishes because ψ either vanishes at infinity or satisfies periodic boundary conditions), and

$$\int_{t_1}^{t_2} \frac{\delta \mathcal{L}}{\delta \dot{\psi}} \frac{\partial}{\partial t}(\delta\psi)\, dt = \frac{\delta \mathcal{L}}{\delta \dot{\psi}} \delta\psi \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t} \left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}} \right) \delta\psi\, dt$$

$$= -\int_{t_1}^{t_2} \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial \dot{\psi}} \delta\psi\, dt, \tag{11.16}$$

we get,

$$\delta S_{21} = \int_{t_1}^{t_2} dt \int_V d^3\mathbf{r} \left[\frac{\partial \mathcal{L}}{\partial \psi} - \text{div} \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)} - \frac{\partial}{\partial t} \frac{\partial \mathcal{L}}{\partial \dot{\psi}} \right] \delta\psi \tag{11.13[2]}$$

According to Hamilton's principle of least (or, stationary) action, this δS_{21} should be zero and that requires, since $\delta\psi$ is arbitrary, the vanishing of the integrand in the square bracket. Thus,

$$\frac{\partial \mathcal{L}}{\partial \psi} - \text{div} \frac{\partial \mathcal{L}}{\partial(\text{grad}\,\psi)} - \frac{\partial}{\partial t} \left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}} \right) = 0. \tag{11.17[1]}$$

(11.17[1]) is the *classical field equation* in terms of the Lagrangian density. It is the analogue, in the case of a field, of the Euler-Lagrange equations (11.3) in classical mechanics. Its similarity to Eq. (11.3) will be even more apparent when expressed in terms of the Lagrangian of the field. Conversion of (11.17[1]) in terms of L is achieved with the help of *functional derivatives*:

Whereas (for a free field) $\mathcal{L}$ is a *function* of ψ and its derivatives, L is a *functional* of ψ and $\dot{\psi}$. The distinction is that while the value of a function at a point $\mathbf{r}$ depends on the value of its arguments (which themselves may be functions) at that point, the value of a functional depends on the values of its arguments over a whole *region* or range. Thus, the value of L in (11.11) depends on the values of

the functions $\psi(\mathbf{r}, t)$ and $\psi(\mathbf{r}, t)$ over the volume V. We may say that the dependence of $\mathcal{L}$ on ψ is *parametric* whereas the dependence of L on ψ is *functional*, and we distinguish the functional dependence by square brackets, thus writing,

$$L = L[\psi(\mathbf{r}, t), \dot{\psi}(\mathbf{r}, t)]. \tag{11.11a}$$

A functional could be regarded as the continuum limit of a function of discrete variables. In the discrete case, L would be a function of the discrete variables $\{\psi_i(t)\}$ and $\{\dot{\psi}_i(t)\}$:

$$L = L(\psi(t), \dot{\psi}(t)), \tag{11.11b}$$

(see Eq. (11.1))[5]. The variation δL in L corresponding to independent variations $\delta\psi_i$ and $\delta\dot{\psi}_i$ is given by,

$$\delta L = \sum_i \left(\frac{\partial L}{\partial \psi_i} \delta\psi_i + \frac{\partial L}{\partial \dot{\psi}_i} \partial\dot{\psi}_i \right). \tag{11.18a}$$

In the continuum limit, this equation is written as

$$\delta L = \sum_i L t_{\delta V_i \to 0} \left(\frac{1}{\delta V_i} \frac{\partial L}{\partial \psi_i} \delta\psi_i + \frac{1}{\delta V_i} \frac{\partial L}{\partial \dot{\psi}_i} \delta\dot{\psi}_i \right) \delta V_i \tag{11.18b}$$

$$= \int_V \left(\frac{\partial L}{\partial \psi} \delta\psi + \frac{\partial L}{\partial \dot{\psi}} \delta\dot{\psi} \right) d^3\mathbf{r}, \tag{11.18c}$$

where,

$$\frac{\partial L}{\partial \psi(\mathbf{r}, t)} = L t_{\delta V_i \to 0} \frac{1}{\delta V_i} \frac{\partial L}{\partial \psi_i(t)} = \frac{\delta L}{\delta \psi}, \tag{11.19a}$$

$$\frac{\partial L}{\partial \dot{\psi}(\mathbf{r}, t)} = L t_{\delta V_i \to 0} \frac{1}{\delta V_i} \frac{\partial L}{\partial \dot{\psi}_i(t)} = \frac{\delta L}{\delta \dot{\psi}}. \tag{11.19b}$$

In (11.18b), δV_i's are interpreted as the volumes of the cells into which the volume V is divided and ψ_i as the value of $\psi(\mathbf{r}, t)$ at the ith cell. The variation of ψ and $\dot{\psi}$ at each cell can be done independently so that, a variation can be defined by either

$$\delta\psi = \delta_{ij}\delta\psi_i, \text{ or } \delta\dot{\psi} = \delta_{ij}\delta\dot{\psi}_i$$

$\frac{\partial L}{\partial \psi}$ and $\frac{\partial L}{\partial \dot{\psi}}$ defined by Eqs. (11.19a, b) are the functional derivatives of L with respect to ψ and $\dot{\psi}$ respectively. We see that $\frac{\partial L}{\partial \psi}$ is essentially the partial derivative of L with respect to the value of ψ at the point $\mathbf{r}$.

Now, from Eqs. (11.13[1]), (11.14) and (11.15), yield,

$$\delta L = \int_V \left[\left(\frac{\partial L}{\partial \psi} - \text{div} \frac{\partial L}{\partial(\text{grad } \psi)} \right) \delta\psi + \frac{\partial L}{\delta \dot{\psi}} \delta\dot{\psi} \right] d^3\mathbf{r}. \tag{11.20}$$

Comparing Eqs. (11.18c) and (11.20), we get,

---

5. It is in view of (11.11b) that grad ψ is not included as an independent argument in (11.11a).

$$\frac{\partial \mathcal{L}}{\partial \psi(\mathbf{r}, t)} = \frac{\partial \mathcal{L}}{\partial \psi(\mathbf{r}\, t)} - \text{div}\, \frac{\partial \mathcal{L}}{\partial(\text{grad}\, \psi(\mathbf{r}, t))}, \tag{11.21a}$$

$$\frac{\partial \mathcal{L}}{\partial \dot{\psi}(\mathbf{r}, t)} = \frac{\partial \mathcal{L}}{\partial \dot{\psi}(\mathbf{r}, t)}. \tag{11.21b}$$

Substituting these in (11.17[1]), the classical field equation takes the form,

$$\frac{\partial \mathcal{L}}{\partial \psi} - \frac{\partial}{\partial t}\left(\frac{\partial \mathcal{L}}{\partial \dot{\psi}}\right) = 0. \tag{11.17[2]}$$

For relativistic fields, it is more convenient to use the covariant form of Eq. (11.17[1]), namely,

$$\frac{\partial \mathcal{L}}{\partial \psi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \psi_{,\mu}} = 0, \tag{11.17[3]}$$

where the summation convention of last chapter is used, and $\psi_{,\mu} \equiv \partial_\mu \psi$. Also, the Lagrangian density could be a function of several independent fields as, for example, in the case of particles with spin. In that case, each field can be varied separately in applying the Hamilton's principle, obtaining an Euler-Lagrange equation for each field:

$$\frac{\partial \mathcal{L}}{\partial \psi^\alpha} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \psi^\alpha_{,\mu}} = 0, (\alpha = 1, 2, \ldots, N). \tag{11.17[4]}$$

Hamiltonian Formulation

In analogy with Eqs. (11.5) and (11.4), the momentum P_i conjugate to the canonical coordinate ψ_i is defined by

$$P_i(t) = \frac{\partial L}{\partial \dot{\psi}_i}; \tag{11.22[1]}$$

and the Hamiltonian of the field by

$$H(t) = \sum_i P_i \dot{\psi}_i - L. \tag{11.23[1]}$$

Going over to the continuum limit, where ψ_i is interpreted as the value of $\psi(\mathbf{r}, t)$ in the ith cell, Eqs. (11.22[1]) and (11.23[1]) become, in view of Eqs. (11.19b) and (11.21b),

$$P_i(t) = \frac{\partial \mathcal{L}}{\partial \dot{\psi}(\mathbf{r}, t)} \delta V_i = \pi(\mathbf{r}, t) \delta V_i, \tag{11.22[2]}$$

and

$$H(t) = \int_V \{\pi(\mathbf{r}, t)\psi(\mathbf{r}, t) - \mathcal{L}(\mathbf{r}, t)\}\, d^3\mathbf{r}$$

$$= \int_V \mathcal{H}(\mathbf{r}, t) d^3\mathbf{r}, \tag{11.23[2]}$$

where,

$$\mathcal{H}(\mathbf{r}, t) = \pi(\mathbf{r}, t)\psi(\mathbf{r}, t) - \mathcal{L}(\mathbf{r}, t), \tag{11.24}$$

is the *Hamiltonian density*. Also,

$$\pi(\mathbf{r}, t) = \frac{\partial L}{\partial \dot{\psi}(\mathbf{r}, t)} = \frac{\delta L}{\delta \dot{\psi}}, \tag{11.25}$$

is referred to as the *conjugate field*. From Eqs. (11.17[2]) and (11.25), it follows that

$$\dot{\pi}(\mathbf{r}, t) = -\frac{\partial L}{\partial \psi(\mathbf{r}, t)}. \tag{11.26}$$

H is easily seen to be a functional of π and ψ. For, from (11.23[2]), (11.18c), (11.25) and (11.26), we have,

$$\delta H = \int_V \{(\delta\pi)\dot{\psi} + \pi(\delta\dot{\psi})\} d^3\mathbf{r} - \int_V \left\{\frac{\delta L}{\delta \psi}\delta\psi + \frac{\delta L}{\delta \dot{\psi}}\delta\dot{\psi}\right\} d^3\mathbf{r}$$

$$= \int_V \{(\delta\pi)\dot{\psi} - \dot{\pi}(\delta\psi)\} d^3\mathbf{r}. \tag{11.27[1]}$$

Therefore,

$$H = H[\psi, \pi], \tag{11.23[3]}$$

and $$\delta H = \int_V \left(\frac{\delta H}{\delta \psi}\delta\psi + \frac{\delta H}{\delta \pi}\delta\pi\right) d^3\mathbf{r}, \tag{11.27[2]}$$

with

$$\frac{\delta H}{\delta \psi} = \frac{\partial \mathcal{H}}{\partial \psi} - \text{div}\, \frac{\partial \mathcal{H}}{\partial(\text{grad}\, \psi)}, \tag{11.28a}$$

$$\frac{\delta H}{\delta \pi} = \frac{\partial \mathcal{H}}{\partial \pi} - \text{div}\, \frac{\partial \mathcal{H}}{\partial(\text{grad}\, \pi)}. \tag{11.28b}$$

Comparing (11.27[1]) and (11.27[2]), we get,

$$\psi(\mathbf{r}, t) = \frac{\delta H}{\delta \pi(\mathbf{r}, t)}; \quad -\dot{\pi}(\mathbf{r}, t) = \frac{\delta H}{\delta \psi(\mathbf{r}, t)}. \tag{11.29}$$

If F is an arbitrary functional of ψ and π, then,

$$\dot{F} = \frac{\partial F}{\partial t} + \int_V \left(\frac{\delta F}{\delta \psi}\dot{\psi} + \frac{\delta F}{\delta \pi}\dot{\pi}\right) d^3\mathbf{r}$$

$$= \frac{\partial F}{\partial t} + [F, H]_{PB}, \tag{11.30}$$

where the Poisson bracket is given by

$$[F, H]_{PB} = \int_V \left(\frac{\delta F}{\delta \psi}\frac{\delta H}{\delta \pi} - \frac{\delta F}{\delta \pi}\frac{\delta H}{\delta \psi}\right) d^3\mathbf{r}. \tag{11.31}$$

According to Eq. (11.18c),

$$\delta\psi(\mathbf{r}, t) = \int_V \frac{\delta\psi(\mathbf{r}, t)}{\delta\psi(\mathbf{r}', t)}\delta\psi(\mathbf{r}', t) d^3\mathbf{r}',$$

But,

$$\delta\psi(\mathbf{r}, t) = \int_V \delta(\mathbf{r} - \mathbf{r}')\delta\psi(\mathbf{r}', t) d^3\mathbf{r}'.$$

so that,

$$\frac{\delta\psi(\mathbf{r}, t)}{\delta\psi(\mathbf{r}', t)} = \delta(\mathbf{r} - \mathbf{r}'). \tag{11.32a}$$

This relation corresponds to $\dfrac{\partial q_i}{\partial q_j} = \delta_{ij}$ of the discrete case. Similarly,

$$\frac{\delta\pi(\mathbf{r}, t)}{\delta\pi(\mathbf{r}', t)} = \delta(\mathbf{r} - \mathbf{r}'). \tag{11.32b}$$

Using (11.32a, b), the following relations are easily obtained:

$$[\psi(\mathbf{r}, t), H]_{PB} = \frac{\delta H}{\delta\pi(\mathbf{r}, t)} = \dot{\psi}, \tag{11.33a}$$

$$[\pi(\mathbf{r}, t), H]_{PB} = -\frac{\delta H}{\delta\pi(\mathbf{r}, t)} = \dot{\pi}, \tag{11.33b}$$

$$[\psi(\mathbf{r}, t), \pi(\mathbf{r}', t)]_{PB} = \delta(\mathbf{r} - \mathbf{r}') \tag{11.34a}$$

$$[\psi(\mathbf{r}, t), \psi(\mathbf{r}', t)]_{PB} = 0 = [\pi(\mathbf{r}, t), \pi(\mathbf{r}', t)]_{PB}. \tag{11.34b}$$

Extension of the above formalism to the case of several *truly* independent (that is, not connected by any constraints) fields is straight forward. In Eqs. (11.25), (11.26) and (11.33a, b) ψ and π are replaced by ψ^α and its conjugate field π^α respectively, while Eqs. (11.24), (11.31) and (11.34a, b) become (in covariant notation),

$$\mathcal{H}(x) = \sum_\alpha \pi^\alpha(x)\psi^\alpha(x) - \mathcal{L}, \tag{11.24a}$$

$$[F, H]_{PB} = \sum_\alpha \int \left(\frac{\delta F}{\delta\psi^\alpha}\frac{\delta H}{\delta\pi^\alpha} - \frac{\delta F}{\delta\pi^\alpha}\frac{\delta H}{\delta\psi^\alpha} \right), \tag{11.31a}$$

$$[\psi^\alpha(\mathbf{r}, t), \pi^\beta(\mathbf{r}', t)]_{PB} = \delta_{\alpha\beta}\delta(\mathbf{r} - \mathbf{r}'), \tag{11.34a'}$$

$$[\psi^\alpha(\mathbf{r}, t), \psi^\beta(\mathbf{r}', t)]_{PB} = 0 = [\pi^\alpha(\mathbf{r}, t), \pi^\beta(\mathbf{r}', t)]. \tag{11.34b'}$$

Quantization of the Field

The transition from a classical field to a quantum field is accomplished by steps identical to those described in the transition from the canonical formulation of classical mechanics to quantum mechanics. The field variables ψ and π are regarded as field operators. The quantal properties of $\hat{\psi}$ and $\hat{\pi}$ (and, through them, those of the physical observables of the field) are specified by prescribing algebraic relations for $\hat{\psi}$ and $\hat{\pi}$. Unlike in the case of quantum mechanics, these relations need not be the ones that follow from replacing the Poisson bracket in (11.34a, b) by the commutator bracket divided by $i\hbar$. However, the equations of motion, which should be considered as operator equations, are obtained by such a replacement of the Poisson brackets by the corresponding commutator brackets.

Instead of specifying algebraic relations for $\hat{\psi}$ and $\hat{\pi}$ directly, one could expand $\hat{\psi}$ in terms of some complete orthonormal set of functions $\{u_k\}$:

$$\hat{\psi}(\mathbf{r}, t) = \sum_k \hat{a}_k(t) u_k(\mathbf{r}), \tag{11.35}$$

and then specify algebraic relations for the coefficient operators $\hat{a}_k$. By definition, u_k satisfies the relationships [cf. Eqs. (2.33) and (2.36)],

$$\int u_k^*(\mathbf{r}) u_l(\mathbf{r}) d^3\mathbf{r} = \delta_{kl}, \tag{11.36a}$$

$$\sum_k u_k(\mathbf{r}) u_k^*(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'). \tag{11.36b}$$

The set $\{u_k\}$ could be the plane wave solutions of the corresponding classical field equation. Eq. (11.35) could be regarded as a Fourier decomposition of the field (see Eqs. (C.1) and (C.1a)) into its normal modes. We will illustrate the procedure in the applications to follow:

11.3 NON-RELATIVISTIC FIELDS

The canonical formulation ensures that the quantized field is (is not) Lorentz covariant if the corresponding classical field is (is not). As our first example we choose a non-relativistic field, namely the *Schrödinger field* the classical field equation of which is

$$i\hbar \frac{\partial \psi}{\partial t} + \frac{\hbar^2}{2m} \nabla^2 \psi - V\psi = 0. \tag{11.37}$$

From the viewpoint of quantum mechanics, Eq. (11.37) is the quantized equation of motion of an ensemble of particles of mass m moving in an external field represented by the potential V. But here we look upon it as a classical field equation. It can be then quantized according to the procedure described in the previous Section. Since it is the second time the equation is being quantized, field quantization of this equation (as also of the relativistic quantum mechanical wave equations) is referred to as *second quantization*. We will see that the second quantization of (11.37) leads to the appearance of the field as an assembly of non-interacting indistinguishable particles (analogous to the normal modes of oscillation of a system of coupled oscillators in classical mechanics).

We start by finding out a Lagrangian density which, when substituted in (11.17[1]), yields Eq. (11.37). We find,

$$\mathcal{L}_S = i\hbar \psi^* \dot{\psi} - \frac{\hbar^2}{2m} \text{ grad } \psi^* . \text{ grad } \psi - V\psi^*\psi. \tag{11.38}$$

The conjugate field $\pi(\mathbf{r}, t)$ is given by [Eq. (11.25)],

$$\pi(\mathbf{r}, t) = \frac{\partial \mathcal{L}_S}{\partial \dot{\psi}} = i\hbar \psi^*(\mathbf{r}, t), \tag{11.39}$$

and the Hamiltonian density $\mathcal{H}$ and Hamiltonian H by,

$$\mathcal{H} = \pi\dot{\psi} - \mathcal{L}_s = \frac{\hbar^2}{2m} \text{ grad } \psi^* . \text{ grad } \psi + V\psi^*\psi$$

$$= -\frac{i\hbar}{2m} \text{ grad } \pi \cdot \text{ grad } \psi - \frac{i}{\hbar} V\pi\psi. \tag{11.40}$$

$$H = \int_v \left(\frac{\hbar^2}{2m} \text{ grad } \psi^* \cdot \text{ grad } \psi + V\psi^*\psi \right) d^3\mathbf{r}. \tag{11.41[1]}$$

Quantization

We expand $\hat{\psi}(\mathbf{r}, t)$ in terms of a complete orthonormal set $\{u_k(\mathbf{r})\}$:

$$\hat{\psi}(\mathbf{r}, t) = \sum_k \hat{a}_k(t) u_k(\mathbf{r}), \tag{11.42a}$$

Also,

$$\hat{\psi}^\dagger(\mathbf{r}, t) = -(i/\hbar)\hat{\pi}(\mathbf{r}, t) = \sum_k \hat{a}_k^\dagger(t) u_k^*(\mathbf{r}). \tag{11.42b}$$

We choose the $\{u_k(\mathbf{r})\}$ to be the energy eigenfunctions of the Hamiltonian of a single particle in the field:

$$\hat{H}_p = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}),$$

with

$$\hat{H}_p u_k(\mathbf{r}) = \epsilon_k u_k(\mathbf{r}). \tag{11.43}$$

(Note that V is assumed to be independent of t).
Quantization is done by postulating suitable algebraic relations for the operators $\hat{a}_k(t)$ and $\hat{a}_k^\dagger(t)$.

System of Bosons

We try the following commutation relations for the Fourier coefficients in (11.42a, b):

$$[\hat{a}_k, \hat{a}_l^\dagger] = \hat{\delta}_{k,l},$$

$$[\hat{a}_k, \hat{a}_l] = \hat{0} = [\hat{a}_k^\dagger, \hat{a}_l^\dagger], \tag{11.44}$$

where all the operators refer to the same time. (11.44) is seen to be identical with (4.76) of Section 4.2A, except that here we have an infinite number of operators $\hat{a}_k$ in place of the single $\hat{a}$ in (4.76). From the result of that Section, we can draw the following conclusions.

The eigenvalue spectrum of the Hermitian operators $\hat{N}_k = \hat{a}_k^\dagger \hat{a}_k$ are the non-negative integers,

$$n_k = 0, 1, 2, \ldots, +\infty. \tag{11.45}$$

A general state vector of the field is given by

$$| n_1, n_2, \ldots n_k, \ldots \rangle = C(\hat{a}_1^\dagger)^{n_1} (\hat{a}_2^\dagger)^{n_2} \ldots (\hat{a}_k^\dagger)^{n_k} \ldots | 0 \rangle. \tag{11.46}$$

where,

$$C = \frac{1}{(n_1!\, n_2! \ldots n_k! \ldots)^{1/2}}, \tag{11.47}$$

and $| 0 \rangle$ is the *vacuum* state defined by

$$\hat{N}_k | 0 \rangle = 0, \text{ for all } k. \tag{11.48}$$

Also,

$$\hat{a}_k | n_1, n_2, \ldots, n_k, \ldots \rangle = \sqrt{n_k} | n_1, n_2, \ldots, (n_k - 1), \ldots \rangle, \tag{11.49a}$$

$$\hat{a}_k^\dagger | n_1, n_2, \ldots, n_k, \ldots \rangle = \sqrt{n_k + 1} | n_1, n_2, \ldots, (n_k + 1), \ldots \rangle, \tag{11.49b}$$

From (11.41), (11.42a, b) and (11.43) we get

$$\hat{H} = \sum_{k,l} \hat{a}_k^\dagger \hat{a}_l \int_V \left(\frac{\hbar^2}{2m} \text{ grad } u_k^* \cdot \text{ grad } u_l + V u_k^* u_l \right) d^3\mathbf{r}$$

$$= \sum_k \hat{N}_k \in_k, \tag{11.41^2}$$

where the orthonormality of the u_k's , [Eq. (11.36a)], has been used. The total energy (the eigenvalues of $\hat{H}$) of the field in a state $| n_1, n_2, \ldots n_k, \ldots \rangle$ is thus,

$$E = \sum_k n_k \in_k, \tag{11.50}$$

As in Section 4.2A, Eqs. (11.45), (11.49a, b) and (11.50) enable us to interpret $\hat{a}_k^\dagger, \hat{a}_k$ and $\hat{N}_k$ respectively as the *creation*, *annihilation* and *particle-number* operators for particles in the state u_k with energy $\in_k$. The vectors (11.46) define an *occupation-number representation* for the system. Since a given particle-state u_k can be occupied by any number of particles, the field represents an assembly of *bosons*.

Problem 11.1: Show that $(\hat{a}_k)^r (\hat{a}_l^\dagger)^s | 0 \rangle = \dfrac{s!}{(s-r)!} (\hat{a}_k^\dagger)^{s-r} | 0 > \delta_{kl}$. Hence deduce Eq. (11.47).

System of Fermions

We have seen that the quantization postulates (11.44) lead to a system of bosons. For a system of fermions, the occupation number n_k should be restricted to 0 and 1. It has been shown by Jordan and Wigner[6] that this condition could be met by

6. Jordan P. and Wigner, E.P. Z. *Physik*, **47**, 631 (1928).

replacing the commutation relations (11.44) by the following anticommutation relations:

$$\{\hat{a}_k, \hat{a}_l^\dagger\} = \{\hat{a}_l^\dagger, \hat{a}_k\} = \hat{\delta}_{kl}, \tag{11.51a}$$

$$\{\hat{a}_k, \hat{a}_l\} = \hat{0} = \{\hat{a}_k^\dagger, \hat{a}_l^\dagger\} \tag{11.51b}$$

Here also, all the operators refer to the same time. (11.51b) requires,

$$\hat{a}_k \hat{a}_k = \hat{0} = \hat{a}_k^\dagger \hat{a}_k^\dagger, \tag{11.52}$$

so that

$$\hat{N}_k^2 = \hat{a}_k^\dagger \hat{a}_k \hat{a}_k^\dagger \hat{a}_k \equiv \hat{a}_k^\dagger (\hat{1} - \hat{a}_k^\dagger \hat{a}_k) \hat{a}_k = \hat{N}_k,$$

or,

$$\hat{N}_k (\hat{N}_k - \hat{1}) = \hat{0}, \tag{11.53}$$

from which it follows that the eigenvalue spectrum of $\hat{N}_k$ is given by,

$$n_k = 0, 1. \tag{11.54}$$

The following results also follow from (11.51a, b) and (11.54):

$$| n_1, n_2, \ldots, n_k, \ldots \rangle = (\hat{a}_1^\dagger)^{n_1} (\hat{a}_2^\dagger)^{n_2} \ldots (\hat{a}_k^\dagger)^{n_k} \ldots | 0 \rangle, \tag{11.55}$$

$$\hat{a}_k | n_1, n_2, \ldots, n_k, \ldots \rangle = (-1)^{S_k} n_k | n_1, n_2, \ldots, (n_k - 1), \ldots \rangle, \tag{11.56a}$$

$$\hat{a}_k^\dagger | n_1, n_2, \ldots, n_k, \ldots \rangle = (-1)^{S_k} (1 - n_k) | n_1, n_2, \ldots, (n_k + 1), \ldots \rangle \tag{11.56b}$$

$$\hat{N}_k | n_1, n_2, \ldots, n_k, \ldots \rangle = n_k | n_1, n_2, \ldots, n_k, \ldots \rangle, \tag{11.56c}$$

where,

$$S_k = \sum_{r=1}^{k-1} n_r. \tag{11.57}$$

Eqs. (11.56a, b) show that an empty state cannot be further emptied and a filled state cannot be further filled. We also see that the annihilation, creation and number operators can be represented by the matrices,

$$(a_k) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \; (a_k^\dagger) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \; (N_k) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{11.58}$$

The Hamiltonian $\hat{H}$ and the total energy E of the field are given in this case by Eqs. (11.41[2]) and (11.50) with n_k restricted to 0 or 1. Also, in the case of bosons as well as fermions, we can define an operator $\hat{N}$ representing the total number of particles by

$$\hat{N} = \sum_k \hat{N}_k. \tag{11.59}$$

From,

$$[\hat{N}, \hat{H}] = \sum_{k,l} [\hat{N}_k, \hat{N}_l] \in_l = \hat{0}, \tag{11.60}$$

it follows that the total number of particles in the field is conserved.

For the case of bosons we have, from (11.42a, b), (11.44) and (11.36b),

$$[\hat{\psi}(\mathbf{r}, t), \hat{\pi}(\mathbf{r}', t)] = i\hbar\hat{\delta}(\mathbf{r}-\mathbf{r}'), \tag{11.44a}$$

$$[\hat{\psi}(\mathbf{r}, t), \hat{\psi}(\mathbf{r}', t)] = 0 = [\hat{\pi}(\mathbf{r}, t), \hat{\pi}(\mathbf{r}', t)]$$

whereas for fermions the corresponding relations are

$$\{\hat{\psi}(\mathbf{r}, t), \hat{\pi}(\mathbf{r}', t)\} = i\hbar\,\hat{\delta}(\mathbf{r}-\mathbf{r}'),$$

$$\{\hat{\psi}(\mathbf{r}, t), \hat{\psi}(\mathbf{r}', t)\} = \hat{0} = \{\hat{\pi}(\mathbf{r}, t), \hat{\pi}(\mathbf{r}', t)\}. \tag{11.51}$$

While (11.44a) resembles the Heisenberg commutation relations (3.12) of quantum mechanics and, thus, could be regarded as the quantum theoretical extension of the classical relations (11.34a, b), Eq. (11.51) has no classical analogue. The implication of this difference could be the following: For a field to be strong enough to be measurable it is necessary to have a large number of particles in the same state so that their contributions to the field are coherent. Therefore, fields that are measurable and have, hence, a classical counterpart should be describable in terms of an assembly of bosons. An example is the electromagnetic field the quanta associated with which have spin 1. It follows that the ψ-field associated with fermions are not measurable though quantities like current density and energy, which are bilinear expressions in $\hat{\psi}$, are measurable.

In terms of $\hat{\psi}$, the particle-number operator $\hat{N}$ (Eq. (11.58)), is given by

$$\hat{N} = \sum_{\mathbf{k}} \hat{a}_k^\dagger a_k = \sum_k \int_V u_k(\mathbf{r})\hat{\psi}^\dagger(\mathbf{r}, t)u_k^*(\mathbf{r}')\hat{\psi}(\mathbf{r}', t)d^3\mathbf{r}d^3\mathbf{r}'$$

$$= \int_V \delta(\mathbf{r}-\mathbf{r}')\hat{\psi}^\dagger(\mathbf{r}, t)\hat{\psi}(\mathbf{r}', t)d^3\mathbf{r}'d^3\mathbf{r}$$

$$= \int_V \hat{\psi}^\dagger(\mathbf{r}, t)\hat{\psi}(\mathbf{r}, t)d^3\mathbf{r}, \tag{11.58a}$$

which is essentially identical to Eq. (1.19). The expression,

$$\hat{a}_k(t) = \int_V u_k^*(\mathbf{r})\hat{\psi}(\mathbf{r}, t)d^3\mathbf{r}, \tag{11.61}$$

which follows from (11.42a) and (11.36a), has been used here.

Commutators and Anticommutators at Unequal Times

The Heisenberg equation of motion for the field operator $\hat{\psi}$ is given, according to Eq. (11.33a), by

$$\frac{\partial\hat{\psi}}{\partial t} \equiv \dot{\hat{\psi}}(\mathbf{r}, t) = \frac{1}{i\hbar}[\hat{\psi}(\mathbf{r}, t), \hat{H}], \tag{11.62a}$$

which is equivalent to

$$\dot{\hat{a}}_k \equiv \frac{d\hat{a}_k}{dt} = \frac{1}{i\hbar}[\hat{a}_k, \hat{H}] = \frac{1}{i\hbar}\sum_l [\hat{a}_k, \hat{N}_l]\ \epsilon_l$$

$$= \frac{1}{i\hbar}\ \epsilon_k\ \hat{a}_k, \tag{11.62b}$$

since

$$[\hat{a}_k, \hat{N}_l] = \delta_{kl}\, \hat{a}_k. \tag{11.63}$$

Thus,

$$\hat{a}_k(t) = \hat{a}_k(0) \exp\,[-(i/\hbar)\, \epsilon_k\, t] \tag{11.64a}$$

and

$$\hat{a}_k^{\dagger}(t) = \hat{a}_k^{(0)} \exp\,[(i/\hbar)\, \epsilon_k\, t]. \tag{11.64b}$$

From these and Eqs. (11.44) and (11.51a, b), we have:

$$[\hat{a}_k(t), \hat{a}_l^{\dagger}(t')]_{\pm} = \delta_{kl} \exp\,[(i/\hbar)\ \epsilon_k\,(t'-t)] \tag{11.65}$$

$$[\hat{a}_k(t), \hat{a}_l(t')]_{\pm} = \hat{0} = [\hat{a}_k^{\dagger}(t), \hat{a}_l^{\dagger}(t')]_{\pm},$$

where

$$[\hat{a}, \hat{b}]_{+} \equiv \{\hat{a}, \hat{b}\}\ ;\ [\hat{a}, \hat{b}]_{-} = [\hat{a}, \hat{b}]. \tag{11.66}$$

Problem 11.2: Verify Eqs. (11.56a-c).

11.4 RELATIVISTIC FIELDS

We have seen that a non-relativistic field can be quantized using either Bose-Einstein statistics (commutation relations for the field operators) or Fermi-Dirac statistics (anticommutation relations). We will see that a given relativistic field can be consistently quantized by using only one of the statistics. This is because of the intimate relationship between a relativistic wave equation and the spin of the particles on the one hand and spins and statistics on the other hand. As examples of relativistic fields we will consider the Klein-Gordon field (which is a scalar field), the Dirac (spinor) field and the electromagnetic (vector) field.

Natural System of Units

It is customary, in relativistic quantum field theory, to use the so-called *natural system of units* (n.s.u.). In this system both $\hbar$ and c are dimensionless and of magnitude unity.

$$\hbar = c = 1. \tag{11.67}$$

From the relations $E = \hbar\omega$ and $\mathbf{p} = \hbar\mathbf{k}$, we see that energy and momentum have then the dimensions of frequency (T^{-1}) and wave number (L^{-1}), respectively. But also, E has the dimension of mc^2 and $\mathbf{p}$ that of mc so that, in natural units, both E and $\mathbf{p}$ should have the dimension of mass (M). Thus, L, T and M^{-1} have the same dimension in the n.s.u. This could have been deduced also from the fact that normally $\hbar$ and c have the dimensions,

$$[\hbar] = ML^2T^{-1};\ [c] = LT^{-1}. \tag{11.68}$$

From the dimensionless fine structure constant [Eq. (10.111)] $e^2/\hbar c \approx 1/137$, it follows that in n.s.u. the electric charge is dimensionless and that the unit of charge e has the magnitude $\approx \sqrt{1/137}$,. In general, a quantity A which has the dimension $[A] = M^{\alpha}L^{\beta}T^{\gamma}$ in the C.G.S. system, has the dimension,

$$[A]_{\text{n.s.u.}} = L^{\beta+\gamma-\alpha} = T^{\beta+\gamma-\alpha} = M^{\alpha-\beta-\gamma}, \tag{11.69}$$

in the natural system. Thus, all quantities can be expressed in terms of just one dimensional quantity which is usually taken to be length (though sometimes time is chosen for certain quantities like energy).

A quantitative relationship between the natural system of units (with length as the dimensional quantity) and the C.G.S. system of units could be established as follows : (11.69) can be obtained by multiplying $[A]$ by $[\hbar]^{\delta}[c]^{\epsilon}$ and then setting the exponents of T and M in the result equal to zero. That is,

$$[A]_{\text{n.s.u.}} = [A]\,[\hbar]^{\delta}[c]^{\epsilon}$$

$$= M^{\alpha+\delta}T^{\gamma-\delta-\epsilon}L^{\beta+2\delta+\epsilon} \tag{11.69^1}$$

Setting $\alpha+\delta = 0 = \gamma-\delta-\epsilon$, we get,

$$\delta = -\alpha;\ \epsilon = \alpha+\gamma,$$

and

$$[A]_{\text{n.s.u.}} = L^{\beta+\gamma-\alpha}.$$

Examples:

(i) Energy: $[E] = ML^2T^{-2}(\alpha = 1, \beta = 2, \gamma = -2)$.

$$\delta = \epsilon = -1,$$

and

$$[E]_{\text{n.s.u.}} = L^{-1},$$

$$\therefore\ E(\text{in cm}^{-1}) = \frac{E(\text{in ergs})}{\hbar c}$$

or,

$$E(\text{in eV}) = E(\text{in cm}^{-1}) \times 197 \times 10^{-7},$$

Thus,

$$1 \text{ fermi}^{-1} = 197 \text{ MeV}$$

$$[\hbar = 1.054 \times 10^{-27} \text{ erg. sec} = 6.582 \times 10^{-16} \text{ eV sec.}]$$

(ii) Electric charge: $[e] = M^{1/2}L^{3/2}T^{-1}$,

$$\delta = \epsilon = -\frac{1}{2}.$$

$$[e]_{\text{n.s.u.}} = L^0 \text{ (dimensionless)}$$

$$e \text{ (in natural units)} = e\frac{(\text{in e.s.u.})}{\sqrt{\hbar c}} \approx \frac{1}{\sqrt{137}}.$$

$$[e \text{ (e.s.u.)} = 4.8 \times 10^{-10}]$$

11.4A. The Klein-Gordon Field

The field equation is (10.6[2]) which, in natural units, takes the form,

$$(\partial_\mu \partial_\mu - m^2)\Phi(x) = 0. \tag{11.70}$$

We assume that Φ is complex. Then, the Lagrangian density is given by

$$\mathcal{L}_{KG}(x) = -\partial_\mu \Phi\, \partial_\mu \Phi^* - m^2 \Phi \Phi^*. \tag{11.71[1]}$$

Here, Φ and Φ^* should be treated as independent fields. Alternatively, Φ and Φ^* could be written as

$$\Phi(x) = \frac{1}{\sqrt{2}}[\Phi_1(x) - i\Phi_2(x)], \tag{11.72a}$$

$$\Phi(x) = \frac{1}{\sqrt{2}}[\Phi_1(x) + i\Phi_2(x)], \tag{11.72b}$$

where $\Phi_1(x)$ and $\Phi_2(x)$ are real fields satisfying Eq. (11.70). In terms of Φ_1 and Φ_2,

$$\mathcal{L}_{KG}(x) = \sum_{r=1}^{2} \left\{ -\frac{1}{2}(\partial_\mu \Phi_r \partial_\mu \Phi_r + m^2 \Phi_r^2) \right\} \tag{11.71[2]}$$

Then,

$$\pi_r(x) = \frac{\partial \mathcal{L}_{KG}}{\partial \dot{\Phi}_r} = i\partial_4 \Phi_r = \dot{\Phi}_r,$$

$$\mathcal{H}_{KG} = \sum_{r=1}^{2} \hat{\pi}_r(x)\dot{\Phi}_r(x) - \mathcal{L}_{KG}(x),$$

and[7]

$$H_{KG} = \sum_{r=1}^{2} \left\{ -\int \partial_4 \Phi_r \partial_4 \Phi_r d^3\mathbf{x} + \frac{1}{2}\int (\partial_\mu \Phi_r \partial_\mu \Phi_r + m^2 \Phi_r^2) d^3\mathbf{x} \right\} \tag{11.73}$$

A plane wave solution of Eq. (11.70) is given by

$$u_{\mathbf{k}}(x) = \frac{1}{\sqrt{V}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} \exp(ikx), \tag{11.74}$$

with

$$kx = k_\mu x_\mu = \mathbf{k} \cdot \mathbf{x} - \omega t, \tag{11.75}$$

$$k^2 \equiv k_\mu k_\mu = \mathbf{k}^2 - \omega_{\mathbf{k}}^2 = -m^2, \tag{11.76a}$$

and

$$\omega_{\mathbf{k}}^2 = \mathbf{k}^2 + m^2. \tag{11.76b}$$

The factor $\frac{1}{\sqrt{2\omega_{\mathbf{k}}}}$ is introduced in (11.74) for interpretational convenience [see Eq. (11.88) below].

---

7. The coordinate vector is hereafter denoted by $\mathbf{x}$ in place of $\mathbf{r}$. Similarly, $d^3\mathbf{x}$ will denote the volume element in the coordinate space.

The field operators $\hat{\Phi}_r$ are expanded in terms of $u_k(x)$:

$$\hat{\Phi}_r(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} \{\hat{a}_r(\mathbf{k}) e^{ikx} + \hat{a}_r^\dagger(\mathbf{k}) e^{-ikx}\}. \tag{11.77}$$

The second term in (11.77), which is the hermitian conjugate of the first term, is needed because the classical field Φ_r is real (so the quantized field is hermitian) unlike the ψ in (11.35).

As quantization postulate, we adopt the commutation relations,

$$[\hat{a}_r(\mathbf{k}), \hat{a}_s^\dagger(\mathbf{k}')] = \hat{\delta}_{rs} \hat{\delta}(\mathbf{k} - \mathbf{k}'),$$

$$[\hat{a}_r(\mathbf{k}), \hat{a}_s(\mathbf{k}')] = \hat{0} = [\hat{a}_r^\dagger(\mathbf{k}), \hat{a}_s^\dagger(\mathbf{k}')]. \tag{11.78}$$

Then, according to the results of Section (11.3), the field will represent a system of bosons. However, there will be two types of particles. In order to see in what respect these particles differ from each other, it is convenient to work in terms of $\hat{\Phi}$ and $\hat{\Phi}^+$ rather than in terms of $\hat{\Phi}_1$ and $\hat{\Phi}_2$. From (11.72a) and (11.77), we have,

$$\hat{\Phi}(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} [\hat{a}(\mathbf{k}) e^{ikx} + \hat{b}^\dagger(\mathbf{k}) e^{-ikx}]$$

$$= \sum_{\mathbf{k}} [\hat{a}(\mathbf{k}) u_{\mathbf{k}}(x) + \hat{b}^\dagger(\mathbf{k}) u_{\mathbf{k}}^*(x)], \tag{11.79}$$

where

$$\hat{a}(\mathbf{k}) = \frac{1}{\sqrt{2}} [\hat{a}_1(\mathbf{k}) - i\hat{a}_2(\mathbf{k})], \tag{11.80a}$$

$$\hat{b}(\mathbf{k}) = \frac{1}{\sqrt{2}} [\hat{a}_1(\mathbf{k}) + i\hat{a}_2(\mathbf{k})]. \tag{11.80b}$$

$\hat{a}, \hat{b}$ and their hermitian conjugates satisfy the commutation rules,

$$[\hat{a}(\mathbf{k}), \hat{a}^+(\mathbf{k}')] = [\hat{b}(\mathbf{k}), \hat{b}^+(\mathbf{k}')] = \hat{\delta}(\mathbf{k} - \mathbf{k}') \tag{11.81}$$

All other combinations vanish.

Define the 4-vector $\hat{S}(x)$ by

$$\hat{S}_\mu(x) = -ie \left(\frac{\partial \hat{L}_{KG}}{\partial \hat{\Phi}, \mu} \hat{\Phi} - \frac{\partial \hat{L}_{KG}}{\partial \hat{\Phi}^\dagger, \mu} \hat{\Phi}^\dagger \right)$$

$$= ie \{(\partial_\mu \hat{\Phi}^\dagger)\hat{\Phi} - (\partial_\mu \hat{\Phi})\hat{\Phi}^\dagger\}. \tag{11.82}$$

Then, from (11.70),

$$\partial_\mu \hat{S}_\mu(x) = iem^2 [\hat{\Phi}^\dagger(x), \hat{\Phi}(x)]$$

$$= \hat{0}, \text{ by Eq. (11.90a) below.} \tag{11.83}$$

$\hat{S}(x)$ can be, thus, interpreted as a 4-current density and S_4/i as the electric charge density. The operator $\hat{Q}$ corresponding to the total charge of the field is given by

$$\hat{Q} = \frac{1}{i} \int \hat{S}_4(x)\, d^3\mathbf{x}$$

$$= e \int \{(\partial_4 \hat{\Phi}^\dagger)\hat{\Phi} - (\partial_4 \hat{\Phi})\hat{\Phi}^\dagger\} d^3\mathbf{x}$$

$$= \frac{e}{2} \sum_{\mathbf{k},\mathbf{k}'} [\{\hat{a}(\mathbf{k}'), \hat{a}^\dagger(\mathbf{k})\} - \{\hat{b}(\mathbf{k}), \hat{b}^\dagger(\mathbf{k}')\}] \tag{11.84a}$$

$$= e \sum_{\mathbf{k}} [\hat{N}_+(\mathbf{k}) - \hat{N}_-(\mathbf{k})], \tag{11.84b}$$

where

$$\hat{N}_+(\mathbf{k}) = \hat{a}^\dagger(\mathbf{k})\, \hat{a}(\mathbf{k}), \tag{11.85}$$

$$\hat{N}_-(\mathbf{k}) = \hat{b}^\dagger(\mathbf{k})\, \hat{b}(\mathbf{k}).$$

In going from (11.84a) to (11.84b), we have used the commutation relations (11.81). Incidentally, we note that $\hat{Q}$ would be null if we use anticommutation relations (Fermi-Dirac statistics).

The total charge of the field is thus,

$$Q = e \sum_{\mathbf{k}} [n_+(\mathbf{k}) - n_-(\mathbf{k})], \tag{11.86}$$

where,

$$n_\pm(\mathbf{k}) = 0, 1, 2, \ldots, +\infty. \tag{11.87}$$

Thus, the operators $\hat{a}(\mathbf{k})$, $a^\dagger(\mathbf{k})$ and $\hat{N}_+(\mathbf{k})$ can be interpreted respectively as the annihilation, creation and number operators for a particle of electric charge $+e$ and momentum $\mathbf{k}$, while $\hat{b}(\mathbf{k})$, $\hat{b}^\dagger(\mathbf{k})$ and $\hat{N}_-(\mathbf{k})$ are similar operators for a particle of charge $-e$. This interpretation can be confirmed by evaluating the field Hamiltonian $\hat{H}$. From (11.73), (11.80) and (11.81), we find,

$$\hat{H}_{KG} = \frac{1}{2} \sum_{\mathbf{k}} \omega_{\mathbf{k}} \sum_{r=1}^{2} \{\hat{a}_r(\mathbf{k}), \hat{a}_r^\dagger(\mathbf{k})\}$$

$$= \frac{1}{2} \sum_{\mathbf{k}} \omega_{\mathbf{k}} [\{\hat{a}(\mathbf{k}), \hat{a}^\dagger(\mathbf{k})\} + \{\hat{b}(\mathbf{k}), \hat{b}^\dagger(\mathbf{k})\}]$$

$$= \sum_{\mathbf{k}} [\hat{N}_+(\mathbf{k}) + \hat{N}_-(\mathbf{k})]\, \omega_{\mathbf{k}} + \hat{H}_0, \tag{11.88}$$

with

$$\hat{H}_0 = \sum_{\mathbf{k}} \omega_{\mathbf{k}} \hat{1}. \tag{11.88a}$$

The field energy is given by

$$E_{KG} = E_+ + E_-, \tag{11.89}$$

with[8]

$$E_\pm = \sum_{\mathbf{k}} \left\{ n_\pm(\mathbf{k}) + \frac{1}{2} \right\} \omega_{\mathbf{k}}. \tag{11.89a}$$

Again we note that the field energy would have been zero had we used anticommutation relations for the creation and annihilation operators. This shows that *the Klein-Gordon field can be consistently quantized only by using Bose-Einstein statistics.*

---

8. Note the similarity of (11.89a) to the corresponding harmonic oscillator expression,

$$E_{h.o} = \sum_{n=0}^{\infty} \left(n + \frac{1}{2} \right) \omega$$

which follows from Eq. (4.94). The field appears as a collection of independent oscillators.

We see from the foregoing analysis that the Klein-Gordon field with a complex $\hat{\Phi}$ represents spinless (since $\hat{\Phi}$ is a scalar) charged particles of charge $+e$ and $-e$. The particle with charge $-e$ is regarded as the antiparticle of the one with charge $+e$. The theory is seen to be symmetric under the interchange of these particles.

We also see from (11.82) that for a hermitian $\hat{\Phi}$ (obtained by setting $\hat{\Phi}_2$ in (11.72) and $\hat{a}_2$ in (11.80) equal to zero), the current and charge densities vanish. But the field Hamiltonian is given by

$$\hat{H} = \sum_{\mathbf{k}} \left\{ \hat{N}(\mathbf{k}) + \frac{1}{2} \right\} \omega_{\mathbf{k}},$$

with $$\hat{N}(\mathbf{k}) = \hat{a}^{\dagger}(\mathbf{k})\, \hat{a}(\mathbf{k}).$$

Therefore, a hermitian (or real) scalar field represents *neutral* spinless particles. In general, thus, the Klein-Gordon equation is the field equation of a spinless particle which exists in three charge states—positive, negative and neutral. Experimentally, the πmesons fit this prediction.

We notice that the field Hamiltonian is positive definite, and the negative energy problem of relativistic quantum mechanics has disappeared.

Problem 11.3: Verify Eqs. (11.84) and (11.88).

Invariant Delta Functions

From (11.79) and (11.81), we have,

$$[\hat{\Phi}(x), \hat{\Phi}^{\dagger}(x')] = \frac{1}{2V} \sum_{\mathbf{k},\mathbf{k}'} \sqrt{\frac{1}{\omega_{\mathbf{k}} \omega_{\mathbf{k}'}}} \{ [\hat{a}(\mathbf{k}), \hat{a}^{\dagger}(\mathbf{k}')] \times \exp [i(kx - k'x')] + [\hat{b}^{\dagger}(\mathbf{k}), \hat{b}(\mathbf{k}')] \exp [-i(kx - k'x')] \}$$

$$= i\Delta(x - x'). \tag{11.90}$$

Here, $$\Delta(x) = \frac{1}{V} \sum_{\mathbf{k}} \frac{1}{\omega_{\mathbf{k}}} \sin kx. \tag{11.91[1]}$$

$\Delta(x)$ is known as the *invariant delta function*, signifying its invariance under Lorentz transformations. The Lorentz invariance of $\Delta(x)$ follows from the fact that $\hat{\Phi}$ and $\hat{\Phi}^{\dagger}$ are Lorentz scalars[9].

The R.H.S. of Eq. (11.91[1]) can be converted into an integral by the following procedure:

Imagine the volume V to be a cube of sides of length L. Then periodic boundary conditions will restrict the components of $\mathbf{k}$ to the values [see Eq. (8.176)],

$$k_i = \frac{2\pi}{L} n_i, (i = 1, 2, 3; n_i = 0, \pm 1, \pm 2, \ldots). \tag{11.92}$$

---

9. For an explicit proof of the Lorentz covariance and for a more detailed account of the invariant delta function, see Ref. 1 chapter 6; Ref. 2, Section 2-3 or Ref. 3, Section 4.6.

A volume element in **k**-space is thus given by

$$d^3\mathbf{k} = dk_1\, dk_2\, dk_3 = \frac{(2\pi)^3}{V} dn, \tag{11.93}$$

where dn represents the number of allowed values of **k** in $d^3\mathbf{k}$. Let us divide the volume in the **k**-space into small cells and let $(d^3\mathbf{k})_j$ represent the jth cell. Then, if $f(\mathbf{k})$ is an arbitrary function of **k**,

$$\sum_{\mathbf{k}\,(\text{over k}-\textit{space})} f(\mathbf{k}) = \sum_j \{f(\mathbf{k})dn\}_{j\text{th cell}}$$

$$= \frac{V}{(2\pi)^3} \sum_j \{f(\mathbf{k})d^3\mathbf{k}\}_j, \text{ by Eq. (11.93)}$$

$$= \frac{V}{(2\pi)^3} \int d^3\mathbf{k}\, f(\mathbf{k}). \tag{11.94}$$

Since the function $f(\mathbf{k})$ is arbitrary, (11.94) implies the correspondence:

$$\frac{1}{V}\sum_{\mathbf{k}} \to \frac{1}{(2\pi)^3}\int d^3\mathbf{k}. \tag{11.94a}$$

Applying (11.94) to (11.91[1]), we get,

$$\Delta(x) = \frac{1}{(2\pi)^3}\int \frac{d^3\mathbf{k}}{\omega_{\mathbf{k}}} \sin kx. \tag{11.91[2]}$$

$$= \Delta^{(+)}(x) + \Delta^{(-)}(x), \tag{11.91[3]}$$

where

$$\Delta^{\pm}(x) = \mp\frac{i}{(2\pi)^3}\int d^3\mathbf{k} \frac{\exp(+ikx)}{2\omega_{\mathbf{k}}}$$

$$= -\Delta^{(\mp)}(-x). \tag{11.95[1]}$$

Now

$$\int d^3\mathbf{k}\frac{e^{ikx}}{2\omega_{\mathbf{k}}} = \int d^3\mathbf{k} \exp(i\mathbf{k}.\mathbf{x}) \int_{\omega>0} d\omega \frac{\exp(-i\omega t)}{2\omega}\delta(\omega_{\mathbf{k}} - \omega)$$

$$= -i \int d^4k\, e^{ikx}\Theta(\omega)\left[\delta(\omega_{\mathbf{k}}^2 - \omega') + \frac{1}{2\omega}\delta(\omega_{\mathbf{k}} + \omega)\right]$$

$$= -i \int d^4k\ e^{ikx}\Theta(\omega)\delta(k^2 + m^2), (\omega_{\mathbf{k}} > 0). \tag{11.96a}$$

Here, Eqs. (11.75), (11.76a, b) and (D.11) are used. Also, $\Theta(\omega)$ is the Heaviside step function (D.7) and $d^4k = d^3\mathbf{k}\, dk_4 = id^3\mathbf{k}\, d\omega$.

Similarly,

$$\int d^3\mathbf{k}\frac{e^{-ikx}}{2\omega_{\mathbf{k}}} = \int d^3\mathbf{k} \exp(-i\mathbf{k}.\mathbf{x}) \int_{\omega>0} d\omega \frac{\exp(+i\omega t)}{2\omega}\delta(\omega_{\mathbf{k}} - \omega)$$

$$= \int d^3\mathbf{k} \exp(i\mathbf{k}.\mathbf{x}) \int_{\omega'<0} d\omega' \frac{\exp(-i\omega' t)}{2\omega'}\delta(\omega_{\mathbf{k}} + \omega')$$

$$= -i \int d^4k \; e^{ikx}\Theta(-\omega)\delta(k^2+m^2), (\omega_{\mathbf{k}} > 0). \tag{11.96b}$$

In going from the first to the second step, the fact that the limits of integration for the components of $\mathbf{k}$ are from $-\infty$ to $+\infty$ is made use of.

Substituting (11.96a, b) in (11.95[1]) and (11.91[3]), we get the covariant forms of the delta functions:

$$\Delta^{\pm}(x) = \mp \frac{1}{(2\pi)^3}\int d^4k e^{ikx}\Theta(\pm\omega)\delta(k^2+m^2), \tag{11.95[2]}$$

$$\Delta(x) = -\frac{1}{(2\pi)^3}\int d^4k e^{ikx} \in (\omega)\delta(k^2+m^2), \tag{11.91[4]}$$

where
$$\in(\omega) = \Theta(\omega) - \Theta(-\omega) = \begin{cases} 1 \text{ for } \omega > 0 \\ -1 \text{ for } \omega < 0 \end{cases} \tag{11.97}$$

From (11.91[2]) and the 3-dimensional generalization of Eq. (D.6a), we get

$$\frac{\partial}{\partial t}\Delta(x-x')_{t'=t} = \frac{-1}{(2\pi)^3}\int d^3\mathbf{k} \exp(-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}'))$$

$$= -\delta(\mathbf{x}-\mathbf{x}') \tag{11.98}$$

The field operator conjugate to $\hat{\Phi}(x)$ is given by

$$\hat{\pi}(x) = \frac{\partial \hat{L}_{KG}}{\partial \dot{\hat{\Phi}}} = \dot{\hat{\Phi}}^{\dagger}. \tag{11.99}$$

From (11.90), (11.99) and (11.98), we derive[10],

$$[\hat{\Phi}(\mathbf{x},t), \hat{\pi}(\mathbf{x}',t)] = i\delta(\mathbf{x}-\mathbf{x}'),$$

$$[\hat{\Phi}(\mathbf{x},t), \hat{\Phi}(\mathbf{x}',t)] = \hat{0} = [\hat{\pi}(\mathbf{x},t), \hat{\pi}(\mathbf{x}',t)]. \tag{11.100}$$

Also, for $t = 0$, $kx = \mathbf{k}\cdot\mathbf{x}$, so that,

$$\Delta(\mathbf{x},0) = \frac{1}{(2\pi)^3}\int d^3\mathbf{k}\frac{\sin\mathbf{k}\cdot\mathbf{x}}{\omega_{\mathbf{k}}} = 0, \tag{11.101[1]}$$

since $\sin\mathbf{k}\cdot\mathbf{x}$ is an odd function of $\mathbf{k}$. Therefore,

$$[\hat{\Phi}(\mathbf{x},t), \hat{\Phi}^{\dagger}(\mathbf{x}',t)] = i\Delta(\mathbf{x}-\mathbf{x}',0) = \hat{0}. \tag{11.90a}$$

Since $\Delta(x)$ is Lorentz invariant, the result (11.101[1]) holds good for any space-like vector x. That is,

$$\Delta(x) = 0, \quad \text{if } x^2 \equiv \mathbf{x}^2 - t^2 > 0. \tag{11.101[2]}$$

A space-like vector connects two events that are separated essentially in space (that is, a Lorentz frame can be found in which the events are simultaneous but take place at two different locations in space). Such events cannot be connected

---

10. The commutation relations (11.100) could also be obtained using the expression,

$$\hat{\pi}(x) = \dot{\hat{\Phi}}^{\dagger} = \frac{i}{\sqrt{V}}\sum_{\mathbf{k}}\sqrt{\frac{\omega_{\mathbf{k}}}{2}}[a^{\dagger}(\mathbf{k})e^{-ikx} - b(\mathbf{k})e^{ikx}],$$

which follows from (11.79).

by a light signal (since $|\mathbf{x}-\mathbf{x}'| > c(t-t')$) and, therefore, cannot interfere with each other. This is known as the *principle of microcausality*. Eqs. (11.90a) and (11.101[2]) are merely expressions of this principle.

Problem 11.4: Verify the commutation relations (11.100).

Problem 11.5: From (11.91[4]) show that $\Delta(x)$ satisfies the Klein-Gordon equation (11.70).

The Positive and the Negative Frequency Parts

Eq. (11.79) can be written as

$$\hat{\Phi}(x) = \hat{\Phi}^{(+)}(x) + \hat{\Phi}^{(-)}(x), \tag{11.79'}$$

where,

$$\hat{\Phi}^{(+)}(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} \hat{a}(\mathbf{k}) e^{ikx}, \tag{11.79'a}$$

$$\hat{\Phi}^{(-)}(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} \hat{b}^{\dagger}(\mathbf{k}) e^{-ikx}. \tag{11.79'b}$$

$\hat{\Phi}^{(+)}$ and $\hat{\Phi}^{(-)}$ are referred to as the *positive frequency part* and the *negative frequency part*, respectively, of $\hat{\Phi}$. $\hat{\Phi}^{(+)}$ contains only annihilation operators while $\hat{\Phi}^{(-)}$ contains only creation operators. Also, in the case of a Hermitian field like the $\hat{\Phi}_r$ in (11.77),

$$\hat{\Phi}^{(+)} = \hat{\Phi}^{(-)\dagger} \text{ (hermitian } \Phi).$$

$\hat{\Phi}^{(+)}$ is called the positive frequency part because it contains the positive exponent: $e^{+ikx} = \exp(i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t))$. Similarly, $\hat{\Phi}^{(-)}$ contains e^{-ikx}. In a theory where negative energies are allowed, the positive exponent goes with positive energy and the negative exponent with negative energy [see Eqs. (11.104) and (11.117) below].

That the positive and the negative frequency parts should be associated respectively with the annihilation and the creation operators could be seen quite generally as follows:

The equation of motion for the operator $\hat{a}(\mathbf{k})e^{ikx}$ is given, according to (11.62a), by

$$\partial_t(\hat{a}(\mathbf{k})e^{ikx}) = i[\hat{H}, \hat{a}(\mathbf{k})e^{ikx}]$$

Taking the matrix element of this equation between field states Φ_1 and Φ_2 of energy E_1 and E_2, we get,

$$-i\omega_{\mathbf{k}} \langle \Phi_2 | \hat{a}(\mathbf{k})e^{ikx} | \Phi_1 \rangle = i(E_2 - E_1) \langle \Phi_2 | \hat{a}(\mathbf{k})e^{ikx} | \Phi_1 \rangle.$$

That is, $E_2 = E_1 - \omega_{\mathbf{k}} < E_1$. Thus, the operator $\hat{a}(\mathbf{k})e^{ikx}$ causes transitions to a state of lower energy and, hence, of lesser number of particles. Similarly, $\hat{\Phi}^{(-)}$ causes transitions to a state of higher energy and larger number of particles.

11.4B. THE DIRAC FIELD

The classical field equation in this case is (10.25[2]) which in natural units reads,

$$(\gamma_\mu \partial_\mu + m)\psi(x) = 0. \tag{11.102}$$

As an appropriate Lagrangian density, we can use either of the following expressions:

$$\mathcal{L}_D(x) = -\overline{\psi}(x)(\gamma_\mu \partial_\mu + m)\psi(x), \tag{11.103a}$$

$$\mathcal{L}'_D(x) = -\frac{1}{2}[\overline{\psi}(x)(\gamma_\mu \partial_\mu + m)\psi(x) + \psi(x)(\gamma_\mu \partial_\mu + m)\overline{\psi}(x)]. \tag{11.103b}$$

We will use the first of these.

As the complete set of plane wave solutions of (11.102) we choose the functions,

$$\left.\begin{aligned} \phi_r(x) &= \frac{1}{\sqrt{V}}\sqrt{\frac{m}{E_p}}u_r(\mathbf{p})e^{ipx} \\ \phi_{r+2}(x) &= \frac{1}{\sqrt{V}}\sqrt{\frac{m}{E_p}}v_r(\mathbf{p})e^{-ipx} \end{aligned}\right\}, (r = 1, 2), \tag{11.104}$$

where

$$u_r(\mathbf{p}) = \left(\frac{E_p + m}{2m}\right)^{1/2}\begin{pmatrix} \xi_r \\ \left\{\dfrac{(\sigma \cdot \mathbf{p})}{E_p + m}\right\}\xi_r \end{pmatrix}, \tag{11.105a}$$

$$v_r(\mathbf{p}) = \left(\frac{E_p + m}{2m}\right)^{1/2}\begin{pmatrix} \left\{\dfrac{(\sigma \cdot \mathbf{p})}{E_p + m}\right\}\xi_r \\ \xi_r \end{pmatrix} \tag{11.105b}$$

with

$$\xi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix};\ \xi_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{11.106}$$

and

$$px = \mathbf{p} \cdot \mathbf{x} - Et. \tag{11.107a}$$

Also,

$$p^2 \equiv p_\mu p_\mu = \mathbf{p}^2 - E^2 = -m^2;\ E_p = \sqrt{\mathbf{p}^2 + m^2}. \tag{11.107b}$$

The normalization factors adopted in (11.104) and (11.105) lead to the following normalizations.

$$\overline{u}_r(\mathbf{p})\,u_s(\mathbf{p}) = -\overline{v}_r(\mathbf{p})v_s(\mathbf{p}) = \delta_{rs}, \tag{11.108a}$$

$$u_r^\dagger(\mathbf{p})\,u_s(\mathbf{p}) = v_r^\dagger(\mathbf{p})v_s(\mathbf{p}) = (E_p/m)\delta_{rs}, \tag{11.108b}$$

$$\overline{u}_r(\mathbf{p})\,v_s(\mathbf{p}) = \overline{v}_r(\mathbf{p})u_s(\mathbf{p}) = 0, \tag{11.108c}$$

$$u_r^\dagger(\mathbf{p})\,v_s(-\mathbf{p}) = v_r^\dagger(\mathbf{p})u_s(-\mathbf{p}) = 0, \tag{11.108d}$$

where $\bar{u}_r(\mathbf{p}) = u_r^\dagger(\mathbf{p})\gamma_4$.

Also,

$$\int_V \phi_\nu^\dagger(x)\phi_{\nu'}(x)d^3\mathbf{x} = \delta_{\nu\nu'}. \tag{11.109}$$

Substituting from (11.104) in (11.102), we find the equations satisfied by the spinors u_r and v_r:

$$(\gamma_\mu p_\mu - im)u_r(\mathbf{p}) = 0, \tag{11.110a}$$

$$(\gamma_\mu p_\mu + im)v_r(\mathbf{p}) = 0. \tag{11.110b}$$

The solutions $\phi_1(x)$ and $\phi_2(x)$ correspond to positive energy ($E = +E_\mathbf{p}$) while $\phi_3(x)$ and $\phi_4(x)$ correspond to negative energy ($E = -E_\mathbf{p}$). These solutions are easily obtained by solving Eqs. (10.52a, b) for the free-particles case. Writing

$$\phi(x) = \xi(\mathbf{p}, E)e^{ipx},$$

$$\chi(x) = \eta(\mathbf{p}, E)e^{ipx}, \tag{11.111}$$

and substituting in (10.52a, b) with $\Phi = 0, \mathbf{A} = \mathbf{0}, c = 1$, we get the coupled equations,

$$(E - m)\xi = (\sigma \cdot \mathbf{p})\eta,$$

$$(E + m)\eta = (\sigma \cdot \mathbf{p})\xi. \tag{11.112}$$

For $E = +E_\mathbf{p}$, we write $\eta = \{(\sigma \cdot \mathbf{p})/(E_\mathbf{p} + m)\}\xi$ and for $E = -E_\mathbf{p}$,

$$\xi = -\{(\sigma \cdot \mathbf{p})/(E_\mathbf{p} + m)\}\eta$$

There are two independent solutions for each sign of energy. These are, after introducing a normalization factor, given by

$$w_r(\mathbf{p}, E_\mathbf{p}) = A\begin{pmatrix} \xi_r \\ B\xi_r \end{pmatrix}, \ (r = 1, 2), \tag{11.113a}$$

$$w_r(\mathbf{p}, -E_\mathbf{p}) = A\begin{pmatrix} -B\xi_r \\ \xi_r \end{pmatrix}, \ (r = 1, 2), \tag{11.113b}$$

where

$$A = \left\{\frac{E_\mathbf{p} + m}{2m}\right\}^{1/2}$$

$$B = \frac{(\sigma \cdot \mathbf{p})}{E_\mathbf{p} + m}, \tag{11.114}$$

and (by choice)

$$\xi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \xi_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{11.115}$$

The u_r and v_r in (11.105a, b) are defined as

$$u_r(\mathbf{p}) = w_r(\mathbf{p}, E_\mathbf{p}), \tag{11.116a}$$

$$v_r(\mathbf{p}) = w_r(-\mathbf{p}, -E_\mathbf{p}), \tag{11.116b}$$

Corresponding to these and with additional normalization factors, we have,

$$\left.\begin{aligned}\phi_r(x) &= \frac{1}{\sqrt{V}}\sqrt{\frac{m}{E_{\mathbf{p}}}}\,(A\xi_r)e^{ipx}\\ \chi_r(x) &= \frac{1}{\sqrt{V}}\sqrt{\frac{m}{E_{\mathbf{p}}}}\,(AB\xi_r)e^{ipx}\end{aligned}\right\},\ (E = E_{\mathbf{p}}), \tag{11.117a}$$

$$\left.\begin{aligned}\phi_{r+2}(x) &= \frac{1}{\sqrt{V}}\sqrt{\frac{m}{E_{\mathbf{p}}}}\,(AB\xi_r)e^{-ipx}\\ \chi_{r+2}(x) &= \frac{1}{\sqrt{V}}\sqrt{\frac{m}{E_{\mathbf{p}}}}\,(A\xi_r)e^{-ipx}\end{aligned}\right\},\ (E = -E_{\mathbf{p}},\ \mathbf{p}\to-\mathbf{p}) \tag{11.117b}$$

Eqs. (11.104) follow from (10.53), (11.117), (11.116) and (11.113).

Quantization

Expanding the field operators $\hat{\psi}(x)$ and $\hat{\bar{\psi}}(x)$ in terms of $\phi_r(x)$, we have,

$$\hat{\psi}(x) = \frac{1}{\sqrt{V}}\sum_{\mathbf{p}}\sqrt{\frac{m}{E_{\mathbf{p}}}}\sum_{r=1}^{2}[\hat{c}_r(\mathbf{p})u_r(\mathbf{p})e^{ipx} + \hat{d}_r^{\dagger}(\mathbf{p})v_r(\mathbf{p})e^{-ipx}]. \tag{11.118a}$$

$$\hat{\bar{\psi}}(x) = \frac{1}{\sqrt{V}}\sum_{\mathbf{p}}\sqrt{\frac{m}{E_{\mathbf{p}}}}\sum_{r=1}^{2}[\hat{c}_r^{\dagger}(\mathbf{p})\bar{u}_r(\mathbf{p})e^{-ipx} + \hat{d}_r(\mathbf{p})\bar{v}_r(\mathbf{p})e^{ipx}]. \tag{11.118b}$$

where summation is over the allowed values of $\mathbf{p}$ [Eq. (11.92)].

Before interpreting $\hat{c}_r, \hat{c}_r^{\dagger}, \hat{d}_r$ and $\hat{d}_r^{\dagger}$ as annihilation and creation operators, we have to make sure that (11.118) leads to sensible expressions for the energy and the charge of the field. We have,

$$\hat{\pi}(x) = \frac{\partial \hat{\mathcal{L}}_D}{\partial \dot{\hat{\psi}}} = -i\frac{\partial \mathcal{L}_D}{\partial \psi_{,4}} = i\hat{\bar{\psi}}(x)\gamma_4 = i\hat{\psi}^{\dagger}. \tag{11.119}$$

$$\begin{aligned}\hat{\mathcal{H}}_D(x) &= \hat{\pi}(x)\,\dot{\hat{\psi}}(x) - \hat{\mathcal{L}}_D(x)\\ &= \hat{\bar{\psi}}(x)\left(\sum_{i=1}^{3}\gamma_i\partial_i + m\right)\hat{\psi}(x)\\ &= -\hat{\bar{\psi}}(x)\gamma_4\partial_4\hat{\psi} = \hat{\psi}^{\dagger}i\partial_t\hat{\psi},\end{aligned} \tag{11.120}$$

where the last step follows from (11.102). Thus,

$$\begin{aligned}\hat{H}_D &= \int \hat{\psi}^{\dagger}(x)i\partial_t\hat{\psi}(x)d^3\mathbf{x}\\ &= \frac{1}{V}\int d^3\mathbf{x}\sum_{\mathbf{p},\mathbf{p}'}\sum_{r,s}\frac{mE_{\mathbf{p}'}}{\sqrt{E_{\mathbf{p}}E_{\mathbf{p}'}}}[\hat{c}_r^{\dagger}(\mathbf{p})\hat{c}_s(\mathbf{p}')\\ &\qquad\times u_r^{\dagger}(\mathbf{p})u_s(\mathbf{p}')\exp(-i(p-p')x) - \hat{d}_r(\mathbf{p})\hat{d}_s^{\dagger}(\mathbf{p}')\\ &\qquad\times v_r^{\dagger}(\mathbf{p})v_s(\mathbf{p}')\exp(i(p-p')x) - \hat{c}_r^{\dagger}(\mathbf{p})\hat{d}_s^{\dagger}(\mathbf{p}')\end{aligned}$$

$$\times u_r^\dagger(\mathbf{p})v_s(\mathbf{p}')\exp(-i(p+p')x) + \hat{d}_r(\mathbf{p})\hat{c}_s(\mathbf{p}')$$

$$\times v_r^\dagger(\mathbf{p})u_s(\mathbf{p}')\exp(i(p+p')x)]$$

$$= \sum_{\mathbf{p}}\sum_{r=1}^{2} E_{\mathbf{p}}[\hat{c}_r^\dagger(\mathbf{p})\hat{c}_r(\mathbf{p}) - \hat{d}_r(p)\hat{d}_r^\dagger(\mathbf{p})]. \tag{11.121^1}$$

Here, Eqs. (11.108b, d) and the relations [(Eq. (D.6a)],

$$\frac{1}{(2\pi)^3}\int \exp(\pm i(p-p')x)d^3\mathbf{x} = \delta(\mathbf{p}-\mathbf{p}'),$$

and [Eq. (11.94a)],

$$\frac{1}{V}\sum_{\mathbf{p}'}\delta(\mathbf{p}-\mathbf{p}') = \frac{1}{(2\pi)^3}\int d^3\mathbf{p}'\delta(\mathbf{p}-\mathbf{p}'),$$

are used.

Similarly (c.f. Eq. (11.82)),

$$\hat{S}_\mu(x) = -ie\left\{\frac{\partial\hat{\mathcal{L}}_D}{\partial\hat{\psi}_{,\mu}}\hat{\psi} - \frac{\partial\hat{\mathcal{L}}_D}{\partial\hat{\psi}_{,\mu}^\dagger}\hat{\psi}^\dagger\right\} = ie\,\hat{\bar{\psi}}(x)\gamma_\mu\hat{\psi}(x) \tag{11.122}$$

$$\hat{Q} = -i\int_V \hat{S}_4(x)d^3\mathbf{x} = e\int_V \hat{\psi}^\dagger\psi d^3\mathbf{x}$$

$$= e\sum_{\mathbf{p}}\sum_{r=1}^{2}[\hat{c}_r^\dagger(\mathbf{p})\hat{c}_r(\mathbf{p}) + \hat{d}_r(\mathbf{p})\hat{d}_r^\dagger(\mathbf{p})]. \tag{11.123^1}$$

From (11.121^1) and (11.123^1) we see that, if we adopt commutation relations for the field operators, then the field Hamiltonian would not be positive definite whereas the total charge would be. On the other hand, adoption of anti-commutation relations results in a positive definite field Hamiltonian (aside from a negative zero-point, or vacuum, energy) though the total charge would be no longer positive-definite. Since a negative total charge is not unphysical like a negative total energy, we conclude that only the latter alternative is acceptable. Thus,

$$\{\hat{c}_r(\mathbf{p}), \hat{c}_s^\dagger(\mathbf{p}')\} = \{\hat{d}_r(\mathbf{p}), \hat{d}_s^\dagger(\mathbf{p}')\} = \delta_{rs}\delta(\mathbf{p}-\mathbf{p}'),$$

$$\{\hat{c}_r, \hat{c}_s\} = \{\hat{c}_r^\dagger, \hat{c}_s^\dagger\} = \{\hat{d}_r, \hat{d}_s\} = \{\hat{d}_r^\dagger, \hat{d}_s^\dagger\} = \hat{0}. \tag{11.124}$$

Then,

$$\hat{H}_D = \sum_{\mathbf{p}}\sum_{r=1}^{2} E_{\mathbf{p}}[\hat{N}_r^-(\mathbf{p}) + \hat{N}_r^+(\mathbf{p})] + \hat{H}_0, \tag{11.121^2}$$

$$\hat{Q} = e\sum_{\mathbf{p}}\sum_{r=1}^{2}[\hat{N}_r^-(\mathbf{p}) - \hat{N}_r^+(\mathbf{p})] + \hat{Q}_0, \tag{11.123^2}$$

and

$$E = \sum_{\mathbf{p}}\sum_{r=1}^{2}[n_r^-(\mathbf{p}) + n_r^+(\mathbf{p})] + E_0, \tag{11.125}$$

$$Q = e\sum_{\mathbf{p}}\sum_{r=1}^{2}[n_r^-(\mathbf{p}) - n_r^+(\mathbf{p})] + Q_0, \tag{11.126}$$

where
$$\hat{N}_r^-(\mathbf{p}) = \hat{c}_r^\dagger(\mathbf{p})\hat{c}_r(\mathbf{p}), \tag{11.127a}$$

$$\hat{N}_r^+(\mathbf{p}) = \hat{d}_r^\dagger(\mathbf{p})\hat{d}_r(\mathbf{p}), \tag{11.127b}$$

$$E_0 = -2\sum_{\mathbf{p}} E_{\mathbf{p}} < 0 \tag{11.128}$$

$$Q_0 = 2e \sum_{\mathbf{p}} 1 < 0 \tag{11.129}$$

$$n_r^\pm = 0, 1. \tag{11.130}$$

We interpret $\hat{c}_r^\dagger(\mathbf{p}), \hat{c}_r(\mathbf{p})$ and $\hat{N}_r^-(\mathbf{p})$ as the creation, annihilation and number operators for a particle of momentum $\mathbf{p}$ and charge e (the electron) and $\hat{d}_r^\dagger(\mathbf{p}), \hat{d}_r(\mathbf{p})$ and $\hat{N}_r^+(\mathbf{p})$ as identical operators for the antiparticle (positron) of charge $-e$. Again, the theory is symmetric under the intercharge of the particle and the antiparticle.

The zero-point energy and charge, in this case, could be interpreted in terms of the hole theory (Section 10.2A).

Spins and Statistics

In the quantization of the Klein-Gordon and the Dirac fields we have a sort of theoretical basis for the empirical correlation that was found to exist (Section 9.2) between the spin of a particle and the statistics obeyed by an ensemble of the particle. For, we find that in order to get a positive definite field Hamiltonian, we have to quantize a Klein-Gordon field using Bose-Einstein statistics and a Dirac field using Fermi-Dirac statistics. In general, it is found that quantization of fields representing integral-spin particles requires Bose-Einstein statistics while those corresponding to half-integral spin particles require Fermi-Dirac statistics.

Covariant Anticommutation Relations

From Eqs. (11.105a), (11.106) and (10.30a, c), we have,

$$\sum_{r=1}^{2} u_r(\mathbf{p})\bar{u}_r(\mathbf{p}) = \frac{1}{2m}\begin{pmatrix} (E_{\mathbf{p}}+m)I & -(\boldsymbol{\sigma}\cdot\mathbf{p}) \\ (\boldsymbol{\sigma}\cdot\mathbf{p}) & -\dfrac{\mathbf{p}^2 I}{E_{\mathbf{p}}+m} \end{pmatrix}$$

$$= \frac{1}{2im}(\gamma_\mu p_\mu + im), \tag{11.131a}$$

and

$$\sum_{r=1}^{2} v_r(\mathbf{p})\bar{v}_r(\mathbf{p}) = \frac{1}{2im}(\gamma_\mu p_\mu - im), \tag{11.131b}$$

where I denotes the 2×2 unit matrix. The operators (11.131a) and (11.131b) respectively project out the positive and the negative energy parts of a plane wave solution of momentum $\mathbf{p}$. This follows from (11.108a, c): We see,

$$\sum_r u_r \bar{u}_r \phi_s(x) = \phi_s, \ (s = 1,2), \tag{11.132a}$$

$$= 0, \ (s = 3,4).$$

$$\sum_r v_r \bar{v}_r \phi_s(x) = 0, \ (s = 1,2),$$

$$= -\phi_s, \ (s = 3,4). \tag{11.132b}$$

Thus, the completeness condition for the spinors is given by,

$$\sum_{r=1}^{2} (u_r \bar{u}_r - v_r \bar{v}_r) = \hat{1}. \tag{11.133}$$

We will make use of these results to obtain the anticommutation relations for the field operators $\hat{\psi}$ and $\hat{\bar{\psi}}$. Each of the four components $\hat{\psi}_\alpha$ ($\alpha = 1,2,3,4$) of $\hat{\psi}$ will have to be treated as independent fields, where, according to (11.118a).

$$\hat{\psi}_\alpha(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{p}} \left(\frac{m}{E_{\mathbf{p}}}\right)^{1/2} \sum_{r=1}^{2} [\hat{c}_r(\mathbf{p}) u_{r\alpha}(\mathbf{p}) e^{ipx} + \hat{d}_r^{\dagger}(\mathbf{p}) v_{r\alpha}(\mathbf{p}) e^{-ipx}] \tag{11.118c}$$

$$= \hat{\psi}_\alpha^{(+)}(x) + \hat{\psi}_\alpha^{(-)}(x). \tag{11.118d}$$

From Eqs. (11.118), (11.124), (11.131) and (11.94a), we get,

$$\{\hat{\psi}_\alpha(x), \hat{\bar{\psi}}_\beta(x')\} = \{\hat{\psi}_\alpha^{(+)}(x), \hat{\bar{\psi}}_\beta^{(-)}(x')\} + \{\hat{\psi}_\alpha^{(-)}(x), \hat{\bar{\psi}}_\beta^{(+)}(x')\} \tag{11.134}$$

$$\{\hat{\psi}_\alpha^{(+)}(x), \hat{\bar{\psi}}_\beta^{(-)}(x')\} = \frac{1}{V} \sum_{\mathbf{p},\mathbf{p}'} \frac{m}{\sqrt{E_{\mathbf{p}} E_{\mathbf{p}'}}} \sum_{r=1}^{2} u_{r\alpha}(\mathbf{p}) \bar{u}_{r\beta}(\mathbf{p}') \delta(\mathbf{p} - \mathbf{p}') e^{i(px - p'x')}$$

$$= \frac{1}{(2\pi)^3} \int_{E = E_{\mathbf{p}}} d^3\mathbf{p} \frac{1}{2iE_{\mathbf{p}}} (\gamma_\mu p_\mu + im)_{\alpha\beta} e^{ip(x-x')}$$

$$= -i(\gamma_\mu \partial_\mu - m)_{\alpha\beta} \cdot \frac{-i}{(2\pi)^3} \int_{E = E_{\mathbf{p}}} d^3\mathbf{p} \cdot \left(\frac{1}{2E_{\mathbf{p}}}\right) e^{ip(x-x')}$$

$$= -iS_{\alpha\beta}^{(+)}(x - x'). \tag{11.134}$$

Similarly,

$$\{\hat{\psi}_\alpha^{(-)}(x), \hat{\bar{\psi}}_\beta^{(+)}(x')\} = \frac{1}{(2\pi)^3} \int_{E = E_{\mathbf{p}}} d^3\mathbf{p} \left(\frac{1}{2iE_{\mathbf{p}}}\right) (\gamma_\mu p_\mu - im)_{\alpha\beta} e^{-ip(x-x')}$$

$$= -i(\gamma_\mu \partial_\mu - m)_{\alpha\beta} \frac{i}{(2\pi)^3} \int_{E = E_{\mathbf{p}}} d^3\mathbf{p} \left(\frac{1}{2E_{\mathbf{p}}}\right) e^{-ip(x-x')}$$

$$= -iS_{\alpha\beta}^{(-)}(x - x'). \tag{11.135}$$

From the definition (11.95[1]) of $\Delta^{(\pm)}$, we have,

$$S_{\alpha\beta}^{(\pm)}(x) = (\gamma_\mu \partial_\mu - m)\Delta^{(\pm)}(x). \tag{11.136}$$

Thus,

$$\left\{\hat{\psi}_\alpha(x),\ \hat{\bar{\psi}}_\beta(x')\right\} = -iS_{\alpha\beta}(x-x'), \tag{11.137a}$$

with

$$S_{\alpha\beta}(x) = S^{(+)}_{\alpha\beta}(x) + S^{(-)}_{\alpha\beta}(x)$$

$$= (\gamma_\mu \partial_\mu - m)\Delta(x). \tag{11.138^1}$$

Similarly,

$$\{\hat{\psi}_\alpha(x), \hat{\psi}_\beta(x')\} = \{\hat{\bar{\psi}}_\alpha(x), \hat{\bar{\psi}}_\beta(x')\} = \hat{0}. \tag{11.137b}$$

Using expressions (11.95^2) and (11.91^4) for $\Delta^{(\pm)}$ and Δ, we find,

$$S^{(\pm)}_{\alpha\beta} = \mp\frac{1}{(2\pi)^3}\int d^4p(\gamma_\mu\partial_\mu - m)_{\alpha\beta}e^{ipx}\Theta(\pm p)\delta(p^2+m^2) \tag{11.136^2}$$

$$S_{\alpha\beta} = -\frac{1}{(2\pi)^3}\int d^4p(\gamma_\mu\partial_\mu - m)_{\alpha\beta}e^{ipx} \in (p)\delta(p^2+m^2). \tag{11.138^2}$$

From these it follows that $S_{\alpha\beta}$ (as also $S^{(\pm)}_{\alpha\beta}$) satisfies the Dirac equation:

$$(\gamma_\mu\partial_\mu + m)_{\alpha\rho}S_{\rho\beta}(x) = \frac{1}{(2\pi)^3}\int d^4p\, e^{ipx} \in (p)(p^2+m^2)\delta(p^2+m^2) = 0. \tag{11.139}$$

Physical observables of the Dirac theory are bilinear in $\hat{\bar{\psi}}$ and $\hat{\psi}$. The anticommutators (11.137a, b) ensure that two such observables commute for space-like separations (Problem 11.7) and, thus, satisfy the principle of micro-causality.

Problem 11.6: Obtain expressions for $\hat{H}_D$ and $\hat{Q}$ using (11.103b) as the Lagrangian density.

Problem 11.7: Show that $[\hat{\bar{\psi}}_\alpha(x)\hat{\psi}_\beta(x),\ \hat{\bar{\psi}}_\rho(x')\hat{\psi}_\sigma(x')] = \hat{0}$, when $(x-x')^2 > 0$.

11.4C. THE ELECTROMAGNETIC FIELD

The Klein-Gordon and the Dirac fields are respectively examples of *scalar* and *spinor* fields. We now take up the case of a *vector* field–the electromagnetic field–which is the only field, among the ones discussed here, that was known to classical physicists. In fact, the technique of field quantization was developed specifically to incorporate the principles of quantum theory into this classical field theory.

The classical field equation is (10.168):

$$\partial_\mu\partial_\mu A_\nu(x) = 0, \quad (\nu = 1,2,3,4). \tag{11.140}$$

which, we see, is the Klein-Gordon equation (11.70) for zero-mass ($m = 0$) particles. However, compared with (11.70) the quantization of (11.140) is made difficult by the following factors:

(i) Eqs. (11.140) is equivalent to Maxwell's equations only if it is combined with the Lorentz condition (10.169) on the four-potential A. But the Lorentz condition implies that all four components of A are not independent. On the other hand, canonical quantization procedure is valid only for independent fields (Remember that in obtaining the Euler-Lagrange equations, variation in each of the 'co-ordinates' is treated as independent).

(ii) Whereas A_1, A_2, A_3 are real, $A_4 (= iV)$[11] is imaginary. This makes it difficult to treat all four A_μ's on the same footing.

(iii) Invariance of the field under Gauge transformation (Eq. (10.176)) necessitates different quantization procedures for different gauges.

In meeting the above difficulties, we choose the Lorentz gauge[12] but ignore, at first, the Lorentz condition and the imaginary character of A_4. That is, we treat all the four A_μ's as independent and Hermitian (in the quantized theory). The difficulties (i) and (ii) above are then overcome by a procedure due to Gupta and Bleuler[13].

In analogy with (11.71[2]), the Lagrangian density is given by

$$\mathcal{L}_{EM}(x) = -\frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu, \tag{11.141}$$

while, in place of (11.74), the plane wave solutions of (11.140) are chosen as:

$$u_\mu(x) = \frac{1}{\sqrt{V}}\frac{1}{\sqrt{2\omega_{\mathbf{k}}}}\varepsilon_\mu(\mathbf{k})e^{ikx}, \ (\mu = 1,2,3,4) \tag{11.142}$$

with
$$k^2 = \mathbf{k}^2 - \omega_{\mathbf{k}}^2 = 0, \tag{11.143a}$$

or,
$$\omega_{\mathbf{k}} = |\mathbf{k}| > 0. \tag{11.143b}$$

Also, the factor $\varepsilon_\mu(\mathbf{k})$ in (11.142) denotes that $u_\mu(x)$ is a *vectorial* plane wave, $\varepsilon_\mu(\mathbf{k})$ being the μcomponent of a 4-vector $\varepsilon(\mathbf{k})$ in the Minkowski space.

Expanding the field operators $\hat{A}_\mu(x)$ in terms of $u_\mu(x)$, we have (c.f. Eq. (11.77)),

$$\hat{A}_\mu(x) = \frac{1}{\sqrt{V}}\sum_{\mathbf{k}}\frac{1}{\sqrt{2\omega_{\mathbf{k}}}}\sum_{r=1}^{4}[\hat{a}_r(\mathbf{k})\epsilon_\mu^{(r)}(\mathbf{k})e^{ikx} + \hat{a}_r^\dagger(\mathbf{k})\epsilon_\mu^{(r)*}(\mathbf{k})e^{-ikx}]. \tag{11.144[1]}$$

---

11. Here, V denotes the scalar potential and should not be confused with the normalization volume V occurring in equations such as (11.144).
12. Quantization of the e.m. field using the radiation gauge [(Eq. (10,169a)] is discussed in Ref. 2, Section 4.2.
13. Reference cited in footnotes 40 and 41, Chapter 10.

Here, $\epsilon^{(r)}(\mathbf{k}), (r = 1,2,3,4)$, form a quartet of complete orthonormal vectors (for each value of $\mathbf{k}$) in the $\mathbf{k}$-space[14]:

$$\epsilon_{\mu}^{(r)*}(\mathbf{k})\,\epsilon_{\mu}^{(s)}(\mathbf{k}') = \delta_{rs}\,\delta_{\mathbf{k},\mathbf{k}'} \tag{11.145a}$$

$$\sum_{r=1}^{4} \epsilon_{\mu}^{(r)}(\mathbf{k})\,\epsilon_{\mu'}^{(r)*}(\mathbf{k}) = \delta_{\mu\mu'} \tag{11.145b}$$

$\epsilon_{\mu}^{(r)}$ is the μ-component of $\epsilon^{(r)}$ (that is, the projection of $\epsilon^{(r)}$ onto the x_{μ}-axis). It follows that,

$$\varepsilon_{\mu}(\mathbf{k}) = \sum_{r=1}^{4} \alpha_r\,\epsilon_{\mu}^{(r)}(\mathbf{k}). \tag{11.146}$$

We make use of the arbitrariness of the vectors $\epsilon^{(r)}$ to choose $\epsilon^{(3)}(\mathbf{k})$ along $\mathbf{k}$ and $\epsilon^{(1)}(\mathbf{k})$ and $\epsilon^{(2)}(\mathbf{k})$ perpendicular to $\mathbf{k}$. We further choose the 3-axis along $\mathbf{k}$. Then,

$$\epsilon_{\mu}^{(r)}(\mathbf{k}) = \delta_{\mu r}\,\epsilon^{(r)}(\mathbf{k}), \tag{11.147}$$

with

$$\left.\begin{aligned}\epsilon^{(l)}(\mathbf{k}) &= (\mathbf{e}^{l}(\mathbf{k}),0),\ l = 1,2,3,\\ \epsilon^{(4)}(k) &= (000,i),\end{aligned}\right\} \tag{11.147a}$$

where $\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3$ are mutually orthogonal unit vectors with

$$\mathbf{e}^3 = \frac{\mathbf{k}}{|\,\mathbf{k}\,|}.$$

With this choice, Eq. (11.144[1]) reduces to,

$$\hat{A}_{\mu}(x) = \frac{1}{\sqrt{V}}\sum_{\mathbf{k}}\frac{1}{\sqrt{2\omega_{\mathbf{k}}}}[\hat{a}_{\mu}(\mathbf{k})\,\epsilon^{(\mu)}(\mathbf{k})e^{ikx} + \hat{a}_{\mu}^{\dagger}(\mathbf{k})\,\epsilon^{(\mu)*}(\mathbf{k})e^{-ikx}]$$

$$= \hat{A}_{\mu}^{(+)}(x) + \hat{A}_{\mu}^{(-)}(x). \tag{11.144[2]}$$

In analogy with (11.78), we postulate the commutation relations,

$$[\hat{a}_{\mu}(\mathbf{k}), \hat{a}_{\nu}^{\dagger}(\mathbf{k}')] = \delta_{\mu\nu}\delta(\mathbf{k}-\mathbf{k}'), \tag{11.148}$$

$$[\hat{a}_{\mu}(\mathbf{k}), \hat{a}_{\nu}(\mathbf{k}')] = [\hat{a}_{\mu}^{\dagger}(\mathbf{k}), \hat{a}_{\nu}^{\dagger}(\mathbf{k}')] = 0.$$

Then, $\hat{a}_{\mu}^{\dagger}(\mathbf{k}), \hat{a}_{\mu}(\mathbf{k})$ and $\hat{a}_{\mu}^{\dagger}(\mathbf{k})\hat{a}_{\mu}(\mathbf{k})$ could be interpreted as the creation, annihilation and the number operators, respectively, of a particle (referred to as photon) of momentum $\mathbf{k}$, *polarization* vector $\epsilon^{(\mu)}$ and energy $\omega_{\mathbf{k}} = |\,\mathbf{k}\,|$. According to definition (11.147), $\epsilon^{(1)}$ and $\epsilon^{(2)}$ represent *transverse* (i.e., perpendicular to the momentum vector $\mathbf{k}$) polarization while $\epsilon^{(3)}$ corresponds to *longitudinal* polarization. A physical interpretation of $\epsilon^{(4)}$ is more difficult, but a photon with polarization along $\epsilon^{(4)}$ is called a *scalar* or *time-like* photon. The fact that there are three polarization states in space indicates that the photons have spin 1 ($S_z = -1, 0, +1$). Also, current density operator (11.82) in this case is,

---

14. Since the system is enclosed in the volume V, $\mathbf{k}$ could be treated as discrete according to Eq. (11.92).

$$\hat{S}_\mu(x) = -ie\left(\frac{\partial \hat{L}_{EM}}{\partial \hat{A}_{\nu,\mu}}\hat{A}_\nu - \frac{\partial \hat{L}_{EM}}{\partial \hat{A}^\dagger_{\nu,\mu}}\hat{A}^\dagger_\nu\right) = \hat{0}. \tag{11.149}$$

So the photons are electrically neutral.

The field $\hat{\pi}_\mu(x)$ conjugate to $\hat{A}_\mu(x)$ is given by,

$$\hat{\pi}_\mu(x) = \frac{\partial \hat{L}_{EM}}{\partial \dot{\hat{A}}_\mu} = -i\frac{\partial \hat{L}_{EM}}{\partial \hat{A}_{\mu,4}} = i\partial_4 \hat{A}_\mu = \hat{A}_\mu(x), \tag{11.150}$$

and the Hamiltonian density $\hat{\mathcal{H}}_{EM}(x)$ by

$$\hat{\mathcal{H}}_{EM}(x) = \hat{\pi}_\mu \dot{\hat{A}}_\mu - \hat{L}_{EM}$$

$$= \frac{1}{2}[\hat{\pi}_\mu\hat{\pi}_\mu + (\nabla\hat{A}_\mu)\cdot(\nabla\hat{A}_\mu)]. \tag{11.151}$$

From (11.151), (11.150), (11.144^2), (11.94a), (D.6a), (11.145a) and (11.143b), we obtain the field Hamiltonian:

$$\hat{H}_{EM} = \int_V \hat{\mathcal{H}}_{EM}(x)d^3\mathbf{x} = \sum_{\mathbf{k}} \frac{1}{2}\{\hat{a}_\mu(\mathbf{k}), \hat{a}^\dagger_\mu(\mathbf{k})\}\,\omega_{\mathbf{k}}$$

$$= \sum_{\mathbf{k}}\sum_{\mu}\left[\hat{N}_\mu(\mathbf{k}) + \frac{1}{2}\right]\omega_{\mathbf{k}}, \tag{11.152}$$

which may be compared with (11.88). The total energy of the field is, thus,

$$E_{EM} = \sum_{\mathbf{k}}\sum_{\mu=1}^{4}\left\{n_\mu(\mathbf{k}) + \frac{1}{2}\right\}\omega_{\mathbf{k}}, \tag{11.153}$$

where $$n_\mu(\mathbf{k}) = 0, 1, 2, \ldots, +\infty. \tag{11.154}$$

The Covariant Commutation Relations

Using Eqs. (11.144^2), (11.145a) and (11.147), it is easily shown that

$$[\hat{A}_\mu(x), \hat{A}_\nu(x')] = i\hat{\delta}_{\mu\nu}D(x - x'), \tag{11.155}$$

with,

$$D(x) = \frac{-i}{V}\sum_{\mathbf{k}}\frac{1}{2\omega_{\mathbf{k}}}(\exp(ikx) - \exp(-ikx)),\ (\omega_{\mathbf{k}} = |\mathbf{k}|)$$

$$= -\frac{1}{(2\pi)^3}\int_V d^3\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{x})\left(\frac{\sin\omega_{\mathbf{k}}t}{\omega_{\mathbf{k}}}\right), \tag{11.156^1}$$

where the result, $\int_V d^3\mathbf{k}\exp(-i\mathbf{k}\cdot\mathbf{x}) = \int_V d^3\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{x})$, has been used. We note that Eq. (11.156^1) is similar to Eq. (11.91^2), the only difference being that whereas $\omega_{\mathbf{k}} = \sqrt{\mathbf{k}^2}$ in the former, $\omega_{\mathbf{k}} = \sqrt{\mathbf{k}^2 + m^2}$ in the latter. It is obvious, then, that

$$D(x) = \underset{m \to 0}{\text{Lt}}\, \Delta(x, m) \tag{11.156^2}$$

$$= \frac{-1}{(2\pi)^3} \int d^4k \exp(ikx) \in (\omega)\delta(k^2), \tag{11.156^3}$$

where $\Delta(x, m) \equiv \Delta(x)$, and the last line follows from (11.91^4). From (11.156^3), we have,

$$\partial_\mu \partial_\mu D(x) = \frac{1}{(2\pi)^3} \int d^4k \exp(ikx) \in (\omega)\, k^2 \delta(k^2) = 0, \tag{11.157}$$

by Eq. (D. 9). Thus, $D(x)$ satisfies the field equation.

Problem 11.8: Verify (11.152).

Problem 11.9: Show that $D(x)$ vanishes except on the light cone ($x^2 \equiv \mathbf{x}^2 - t^2 = 0$).

Problem 11.10: Establish the following equal-time commutation relations:

$$[\hat{A}_\mu(x), \hat{\pi}_\nu(x')]_{t'=t} = i\delta_{\mu\nu}\delta(\mathbf{x} - \mathbf{x}'),$$

$$[\hat{A}_\mu(x), \hat{A}_\nu(x')]_{t'=t} = [\hat{\pi}_\mu(x), \hat{\pi}_\nu(x')]_{t'=t} = \hat{0}.$$

The Gupta-Bleuler Formalism

As already stated, the foregoing quantization procedure suffers from two drawbacks. Firstly, it ignores the Lorentz condition,

$$\partial_\mu \hat{A}_\mu(x) = \hat{0}, \tag{11.158}$$

and secondly it assumes all the four field operators $\hat{A}_\mu$ to be Hermitian whereas $\hat{A}_4$ should be antihermitian. In the Gupta-Bleuler formalism being discussed below, these problems are solved by retaining the Hermitian nature of the $\hat{A}_\mu$'s but modifying the usual definition (2.9) of the scalar product. It is further emphasized that correspondence with classical theory requires only the *expectation values* of operators, rather than the operators themselves, to obey the classical equations.

Thus, the definition, $||\Phi||^2 = \langle \Phi \mid \Phi \rangle$, of the norm of Φ is replaced by

$$||\Phi||^2 = \langle \Phi \mid \hat{\eta} \mid \Phi \rangle. \tag{11.159}$$

$\hat{\eta}$ is called the *metric operator*. We will require the norm to be real. Then (see Eq. (2.103)),

$$\hat{\eta}^\dagger = \hat{\eta}. \tag{11.160a}$$

The norm of a vector needs no longer be positive-definite. The vectors divide into three classes depending on the norm being positive, negative or zero. Only for the first of these classes the probability interpretation will go through, so that we will want all physically significant states to be restricted to this class.

With the modified norm (11.159), the expectation value of an operator $\hat{A}$ becomes,

$$\langle \hat{A} \rangle = \langle \Phi \mid \hat{\eta}\, \hat{A} \mid \Phi \rangle. \tag{11.161}$$

$\langle \hat{A} \rangle$ is real for a Hermitian operator $\hat{A}$ if $[\hat{\eta}, \hat{A}] = \hat{0}$ and is imaginary if $\{\hat{\eta}, \hat{A}\} = \hat{0}$. Therefore, $\hat{\eta}$ should be so chosen as to satisfy the conditions,

$$\left.\begin{aligned} [\hat{\eta}, \hat{A}_j(x)] &= \hat{0},\ j = 1,2,3, \\ \{\hat{\eta}, A_4(x)\} &= \hat{0}. \end{aligned}\right\} \tag{11.160b}$$

Also,

$$[\hat{\eta}^2, \hat{A}_j] = \hat{\eta}[\hat{\eta}, \hat{A}_j] + [\hat{\eta}, \hat{A}_j]\hat{\eta} = \hat{0},$$

$$[\hat{\eta}^2, \hat{A}_4] = \hat{\eta}\{\hat{\eta}, \hat{A}_4\} - \{\hat{\eta}, \hat{A}_4\}\hat{\eta} = \hat{0}. \tag{11.162}$$

Thus, $\hat{\eta}^2$ is a c-number (one that commutes with all operators of the field) and can, therefore, be chosen so that

$$\hat{\eta}^2 \equiv \hat{\eta}\hat{\eta}^\dagger = \hat{1}. \tag{11.160c}$$

Determination of $\hat{\eta}$

Eqs. (11.160a-c) are sufficient to determine $\hat{\eta}$. From (11.144[2]) and (11.160b) we have,

$$[\hat{\eta}, \hat{a}_r(\mathbf{k})] = \hat{0},\ (r = 1,2,3). \tag{11.162a}$$

$$\{\hat{\eta}, \hat{a}_4(\mathbf{k})\} = \hat{0}.$$

For simplicity, we assume initially that the state Φ represents photons of momentum $\mathbf{k}$ only. Then, in the occupation number representation (Eq. (11.46)),

$$\Phi = \mid n_1, n_2, n_3, n_4 \rangle, n_\mu \equiv n_\mu(\mathbf{k}),$$

and (Eq. (11.49)),

$$\begin{aligned} &\langle n_1', n_2', n_3', n_4' \mid \hat{a}_r(\mathbf{k}) \mid n_1, n_2, n_3, n_4 \rangle \\ &= \langle n_1, n_2, n_3, n_4 \mid \hat{a}_r^\dagger(\mathbf{k}) \mid n_1', n_2', n_3', n_4' \rangle \\ &= \left\{\sqrt{n_r(\mathbf{k})}\, \delta_{n_r, n_r'+1}\right\} \prod_{i \neq r} \delta_{n_i n_i'}. \end{aligned} \tag{11.163}$$

From (11.162) and (11.163), we get,

$$\langle \ldots, n_r' \ldots \mid \hat{n} \mid \ldots n_r \ldots \rangle = \langle \ldots, n_r + 1, \ldots \mid \hat{\eta} \mid \ldots n_r + 1, \ldots \rangle,\ (r = 1,2,3)$$

(11.164a)

$$< \ldots, n_4 \mid \hat{\eta} \mid \ldots, n_4 > = - < \ldots, n_4 + 1 \mid \hat{\eta} \mid \ldots, n_4 + 1 > . \tag{11.164b}$$

That is, η (the matrix) is diagonal.

Also, by Eq. (11.160c),

$$\mid \langle n_1, n_2, n_3, n_4 \mid \hat{\eta} \mid n_1, n_2, n_3, n_4 \rangle \mid^2 = 1. \tag{11.164c}$$

Eqs. (11.164a-c) show conclusively that

$$\langle n_1', n_2', n_3', n_4' \mid \hat{\eta} \mid n_1, n_2, n_3, n_4 \rangle = (-1)^{n_4} \prod_{r=1}^{4} \delta_{n_r, n_r'}.$$

(11.165[1])

If Φ corresponds to photons of several **k**-values, then,

$$\langle \Phi \mid \hat{\eta} \mid \Phi \rangle = (-1)^{S_4} \prod_{\mathbf{k}} \prod_{r=1}^{4} \delta_{n_r(\mathbf{k}), n_r'(\mathbf{k})}, \qquad (11.165^2)$$

where $S_4 = \sum_{\mathbf{k}} n_4(\mathbf{k})$.

The Lorentz Condition

In the classical theory, the Lorentz condition is an expression of the observed *transversality of the electromagnetic field. As already pointed out, for the quantum* theory to have the correct classical limit, it is sufficient if the expectation values obey the classical equations. Therefore, in place of (11.158), we require,

$$\langle \partial_\mu \hat{A}_\mu \rangle \equiv \langle \Phi \mid \hat{\eta} \partial_\mu \hat{A}_\mu \mid \Phi \rangle = 0. \qquad (11.166)$$

Now,

$$\langle \Phi \mid \hat{\eta} \partial_\mu \hat{A}_\mu \mid \Phi \rangle \equiv (\Phi, \hat{\eta} \partial_\mu \hat{A}_\mu \Phi) = (\Phi, \hat{\eta} \partial_\mu \hat{A}_\mu^{(+)} \Phi) + (\Phi, \hat{\eta} \partial_\mu \hat{A}_\mu^{(-)} \Phi), \qquad (11.167a)$$

and using (2.54) and (11.160b),

$$(\Phi, \hat{\eta} \partial_\mu \hat{A}_\mu^{(-)} \Phi) = (\hat{\eta} \partial_\mu \hat{A}_\mu^{(+)} \Phi, \Phi). \qquad (11.167b)$$

Here $\hat{A}_\mu^{(+)}$ and $\hat{A}_\mu^{(-)}$ are respectively the positive and the negative frequency parts of $\hat{A}_\mu$. Thus, (11.166) is equivalent to the subsidiary condition,

$$\partial_\mu \hat{A}_\mu^{(+)} \Phi = \mathbf{0}. \qquad (11.168^1)$$

But, $\hat{A}_\mu^{(+)}(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} \hat{a}_\mu(\mathbf{k}) \in^{(\mu)} (\mathbf{k}) e^{ikx}$,

so that,

$$\partial_\mu \hat{A}_\mu^{(+)}(x) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \frac{i}{\sqrt{2\omega_{\mathbf{k}}}} \hat{a}_\mu \in^{(\mu)} k_\mu e^{ikx}$$

$$= \frac{i}{\sqrt{2V}} \sum_{\mathbf{k}} \sqrt{\omega_{\mathbf{k}}} \{ \hat{a}_3(\mathbf{k}) - \hat{a}_4(\mathbf{k}) \} e^{ikx},$$

where, the fact that $k_1 = k_2 = 0$, and the relations, $\in^{(3)} k_3 = |\mathbf{k}| = \omega_{\mathbf{k}}$; $\in^{(4)} k_4 = -\omega_{\mathbf{k}}$, are made use of. Thus, Eq. (11.168[1]) could be written as

$$[\hat{a}_3(\mathbf{k}) - \hat{a}_4(\mathbf{k})] \Phi = 0. \qquad (11.168^2)$$

In this form, Eq. (11.166) is seen to be a restriction on the *allowed states* Φ of the field rather than a condition on the operators $\hat{A}_\mu$. We see that the restriction is on the combination of the longitudinal and the scalar photons, and does not affect the transverse photons. In fact, based on a state with only transverse photons, we can

construct a series of states with the same set of transverse photons but different combinations of longitudinal and scalar photons in accordance with Eq. (11.168^2). It can be shown that the different states so constructed are related by gauge transformations. For a free field, the gauge can be so chosen that only transverse photons are present. The longitudinal and the scalar photons are, however, required to represent the Coulomb interaction between charges.

Problem 11.11: Show that the state,

$$\Phi_{\mathbf{k}}^{(N)} = \frac{1}{\sqrt{N!}} [\hat{a}_3(\mathbf{k}) + \hat{a}_4(\mathbf{k})]^N \Phi_{\mathbf{k}}^{LS}(0,0),$$

satisfies condition (11.168^2), where $\Phi_{\mathbf{k}}^{LS}(n_3, n_4)$ denotes a state with n_3 longitudinal and n_4 scalar photons.

11.5 INTERACTING FIELDS

In the preceding Section, we have considered the quantum theory of free fields. Applications to physical problems would, however, require the consideration of interacting fields. Interaction between two fields, in quantum field theory, is viewed as a *coupling* between the fields by means of a *coupling constant*. For example, the interaction between electrons and photons is described in terms of a coupling between the Dirac and the electromagnetic fields, the electric charge (of the electron) acting as the coupling constant [Eq. (11.193) below].

Unlike the case of free fields where the eigenvalue problem is exactly soluble, the problem of coupled fields turns out to be much more complicated and difficult for solution. Only approximation methods have been devised for tackling the problem. An account either of the various difficulties encountered in the case of the coupled fields or of the methods devised to overcome these is beyond the scope of this book[15]. Instead, we will restrict ourselves to a qualitative discussion of certain concepts and techniques employed in the treatment of coupled fields, using the coupling between the Dirac field and an applied (external) electromagnetic field for illustration. We start with the technique of Feynman diagrams.

Feynman Diagrams

Feynman diagrams are graphical representations of certain scattering processes, introduced for the first time by Feynman[16] in connection with the interaction of electromagnetic field with charges. Feynman introduced these on the basis of intuitive arguments. Dyson has given a mathematical interpretation to these diagrams. The Feynman diagrams have proved to be a very useful aid in the picturisation and interpretation of the complex mathematics underlying scattering theory.

---

15. Such accounts could be found in some of the references given at the end of the chapter.
16. R.P. Feynman, *Phys. Rev.* **76**, 749 and 769 (1949).

Consider the scattering of an electron by a potential (representing the electromagnetic field). The electron can be represented by a *world line* in space-time. We will denote the space axis by x and time axis (taken to be perpendicular to the space-axis) by t (see Fig. 11.1). A solid line parallel to the t-axis, as in Fig. 11.1(a), denotes an electron[17] at rest while an inclined line as in (b) represents an electron with a uniform velocity. In these diagrams, if we dispense with the *explicit* drawing of the space and time axes, we have the Feynman diagrams of an electron at rest and an electron with a uniform velocity respectively. Thus, in a Feynman diagram time runs vertically upward[18] and the space coordinate

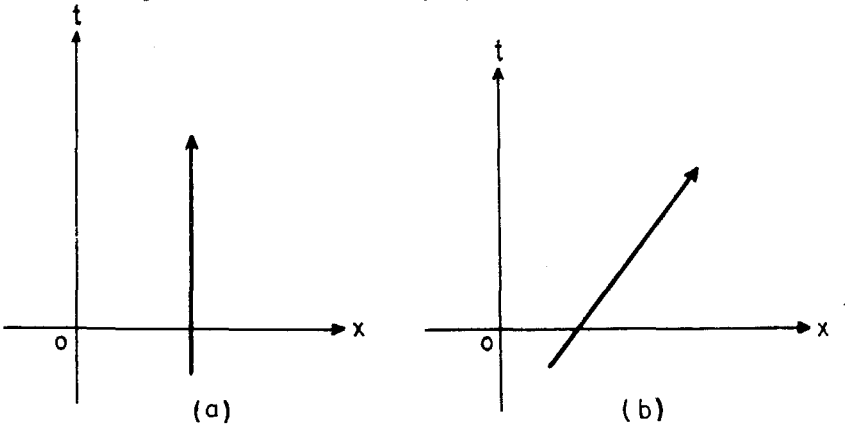


Fig. 11.1. (a) Electron at rest, (b) Electron with a uniform velocity.

increases from left to right. According to this scheme, Fig. 11.2(a) represents the scattering of an electron. The electron starts from[19] $\mathbf{x}$ at time t, gets scattered by the potential at $\mathbf{x}_1$ at time $t_1 > t$, and reaches $\mathbf{x}'$ at $t' > t_1$. Fig. 11.2(b) represents another way of looking at the same process (as far as the initial and the final states of the electron, which only are observed, are concerned). An electron-positron pair is created at $\mathbf{x}_2$ at time $t_2 > t$, the positron proceeds to $\mathbf{x}_1$ where it annihilates with the original electron at time $t_1 > t_2$. The electron of the pair proceeds to $\mathbf{x}'$. Now we notice one thing: if we reverse the direction of the arrow on the world line of the positron between $\mathbf{x}_1$ and $\mathbf{x}_2$ (Fig. 11.2 (c)), then the whole process is represented by a single world line, namely that of the electron which, however, runs *backward* in time between $\mathbf{x}_1$ and $\mathbf{x}_2$. Thus, positron could be described as an electron running backward in time. The world line itself could be looked upon as representing the progress of an *electric charge* which can go forward as well as backward in time, the continuity of the line ensuring the conservation (indestructibility) of the charge.

---

17. Eventhough we specifically refer to the electron, the electron could be replaced by any fermion as far as the diagram is concerned.
18. True only for the so-called *external* lines defined later. In the case of *internal* lines, time may run upward as well as downward.
19. We use bold $\mathbf{x}$ to denote the space co-ordinates alone while x is used to denote both $\mathbf{x}$ and t together.

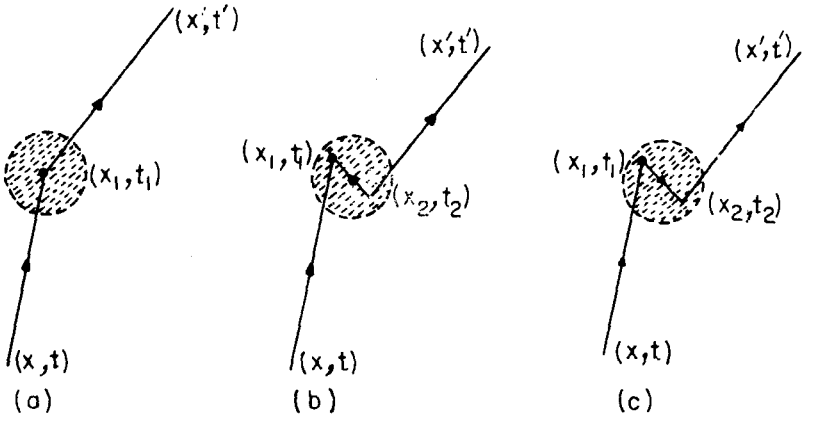


Fig. 11.2. Scattering of an electron by a potential.

From the foregoing we conclude that we can represent an electron (a fermion) graphically by a *solid* line pointing forward in time and a positron (antifermion) by one pointing backward in time.

Let us now have a closer look at the scattering process. We will later see that the interaction between the electromagnetic field and the electric charges is described, in quantum field theory, in terms of emission and absorption of photons by the electric charges. A photon is represented by a wavy[20] line in the Feynman diagram. Thus, the Feynman diagram corresponding to the scattering of an electron by electromagnetic field is as shown in Fig. 11.3(a). An electron of momentum $\mathbf{p}$ absorbs a photon (labelled γ) of momentum $\mathbf{k}$ at x and has its momentum altered to $\mathbf{p}' = \mathbf{p} + \mathbf{k}$. Similarly, Fig. 11.3(b) represents positron

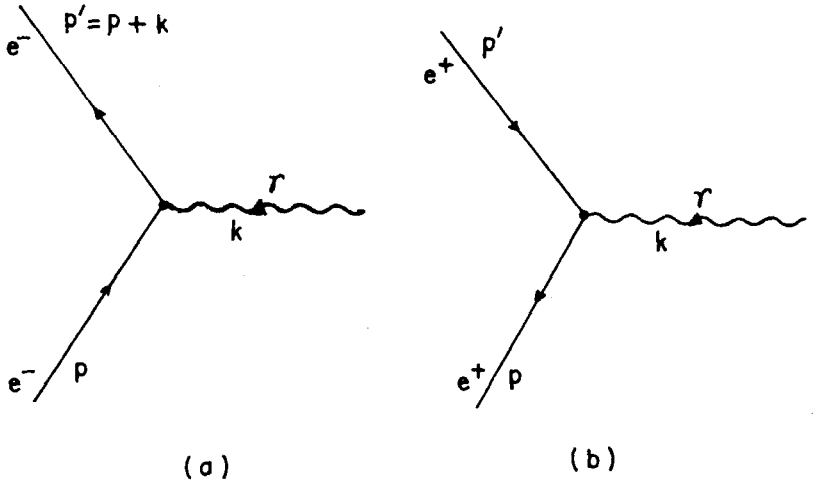


Fig. 11.3. (a) Electron scattering, (b) Positron scattering.

20 Other bosons are represented by *dotted* lines.

scattering. The point x in the diagram is called a *vertex*. The conservation laws (relating to momentum, energy, etc.) should be obeyed at the vertex. For this reason the arrow on the photon line may be omitted since that can be determined from the conservation of momentum and the direction of arrows on the fermion lines. The following points also may be noted:

*FD*1. A solid line ending at the vertex represents either annihilation (or absorption) of an electron or creation (emission) of a positron. Similarly, a solid line beginning at the vertex denotes either creation of an electron or annihilation of a positron.

*FD*2. The number of fermion lines ending at the vertex is equal to the number of fermion lines beginning at the vertex (conservation of fermion number). Following a fermion line through the vertex, the arrow always points in the same sense.

The above rules hold good quite generally for all Feynman diagrams. As a further example, we give in Fig. 11.4 the Feynman diagrams corresponding to single-quantum pair creation [(a)] and pair annihilation [(b)].

Diagrams of the type shown in Figs. 11.3 and 11.4 are known as *basic vertex parts*. These do not, necessarily, correspond to actual physical processes. For example, the single-quantum pair creation (or annihilation) shown in Fig. 11.4 violates the conservation of linear momentum. (This is seen by going over to the centre of mass frame of the electron-positron pair, where the photon momentum does not vanish.)

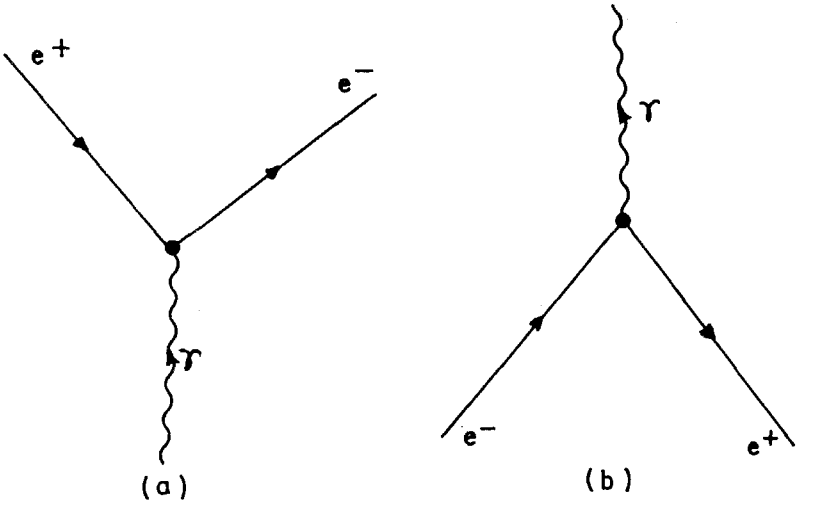


Fig. 11.4. (a) Pair creation, (b) Pair annihilation.

Diagrams corresponding to actual physical processes can be obtained by combining two or more basic vertex parts. Figs. 11.5 and 11.6 illustrate such a procedure.

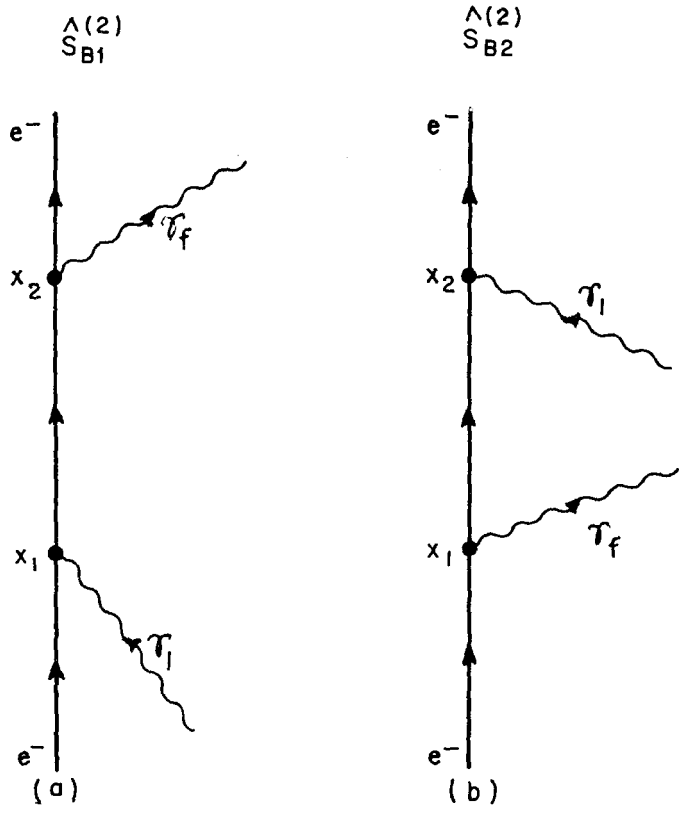
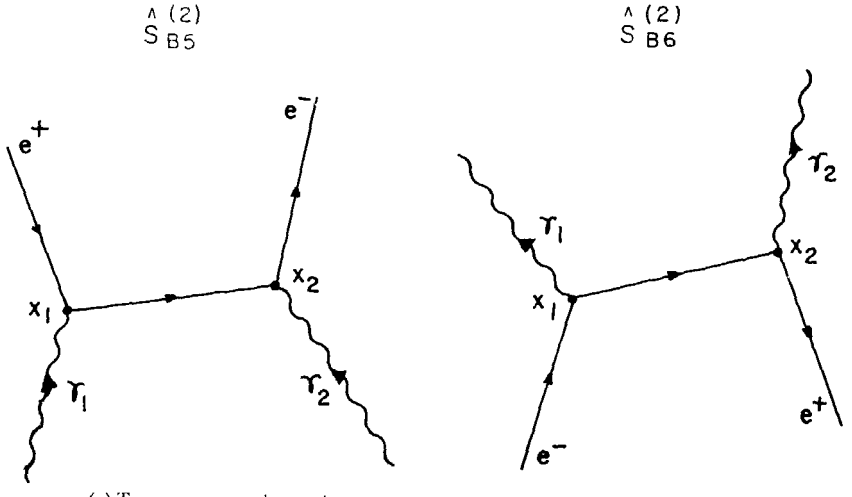


Fig. 11.5. Compton scattering by electrons



(a) Two-quantum pair creation

(b) Two-quantum pair annihilation

Fig. 11.6

Fig. 11.5(a) represents Compton scattering by electrons. If $t_1 < t_2$, the electron absorbs the initial photon at x_1, is propagated as a virtual electron from x_1 to x_2 and emits the final photon at x_2. If $t_1 > t_2$, the same diagram represents the creation of an electron-positron pair at x_2 with emission of a photon, propagation of the (virtual) positron to x_1 where it annihilates itself with the incident electron with the absorption of a photon. Fig. 11.5(b) represents the same phenomenon with initial and final photons exchanged.

Figs. 11.6(a) and 11.6(b) represent respectively two-quantum pair-creation and annihilation. In 11.6(a), an electron-positron pair is created at $x_1 (t_1 < t_2)$ with the absorption of γ_1, the electron is propagated to x_2 where it absorbs γ_2. If $t_1 > t_2$, the electron-positron pair is created at x_2, the positron is propagated to x_1 where it is absorbed along with γ_1 and is remitted. A similar interpretation applies to Fig. 11.6(b).

Problem 11.12: Draw the Feynman diagrams corresponding to (a) pair annihilation in the field of a proton and (b) the decay of a neutron into a proton, an electron and an antineutrino.

Normal Products

An operator $\hat{O}$ which is a product of creation and annihilation operators is a normal product if all annihilation operators stand to the right of all creation operators. Thus,

$$\text{Normal product of } \hat{c}_r \hat{c}_s^\dagger \equiv :\hat{c}_r \hat{c}_s^\dagger: = \pm \hat{c}_s^\dagger \hat{c}_r, \tag{11.169a}$$

where the upper sign is to be chosen when $\hat{c}_s^\dagger$ and $\hat{c}_r$ are boson operators (hence commuting) and the lower sign when they are fermion operators (anticommuting). In general,

$$:\hat{A}\hat{B}\hat{C}\ldots\hat{L}: = (-1)^N \hat{Q}\hat{R}\ldots\hat{W}, \tag{11.169b}$$

where $\hat{Q}, \hat{R}, \ldots \hat{W}$ are operators $\hat{A}, \hat{B}, \ldots \hat{L}$ re-ordered so that all the annihilation operators stand to the right of all creation operators, and N is the number of interchanges of pairs of fermion operators required to accomplish the reordering.

Obviously,

$$:(\hat{A}\hat{B}): = \pm :(\hat{B}\hat{A}):, \tag{11.169c}$$

the upper (lower) sign applying for bosons (fermions).

Normal product obeys the *distributive law*:

$$:(\hat{A}\hat{B} + \hat{C}\hat{D}): = :\hat{A}\hat{B}: + :\hat{C}\hat{D}:$$

A very important property of a normal product is that its vacuum expectation value is zero. For, either the right-most member of the normal product is an annihilation operator in which case, $:(\hat{A}\hat{B}\ldots): |\Phi_0\rangle = 0$, so that $\langle\Phi_0| :(\hat{A}\hat{B}\ldots): |\Phi_0\rangle = 0$, or else, the normal product contains only creation operators in

which case $:(\hat{A}\hat{B}\ldots):|\Phi_0>$ is orthogonal to the vacuum state $|\Phi_0>$. This property of the normal product is responsible for its usefulness. For, if we can express an operator $\hat{0}$ which is a product of creation and annihilation operators in terms of normal products, then the vacuum expectation value of $\hat{0}$ can be easily evaluated since it is the vacuum expectation value of what is left of $\hat{0}$ after removing the normal products. Thus, in the case of a product of two operators, we can write,

$$\hat{A}\,\hat{B} = :\hat{A}\hat{B}: + <\Phi_0|\hat{A}\hat{B}|\Phi_0>. \tag{11.170}$$

Dyson Chronological Product (P)

For two operators $\hat{A}(x)$ and $\hat{B}(x')$, the Dyson chronological product (DCP) is defined by

$$P\{\hat{A}(x)\hat{B}(x')\} = \begin{cases}\hat{A}(x)\,\hat{B}(x'), & \text{if } t > t'\\ \hat{B}(x')\hat{A}(x), & \text{if } t < t'\end{cases} \tag{11.171^1}$$

That is, in DCP the operators occur in chronological order with the time running from right to left.

Using the function $\epsilon(x)$ defined by Eq. (11.97), we have (with $x_0 \equiv t$),

$$P(\hat{A}(x)\hat{B}(x')\} = \frac{1}{2}[1+\epsilon(x_0-x'_0)]\hat{A}(x)\hat{B}(x') + \frac{1}{2}[1-\epsilon(x_0-x'_0)]\hat{B}(x')\hat{A}(x) \tag{11.171^2}$$

$$= P\{\hat{B}(x')\hat{A}(x)\}.$$

Wick's Chronological Product (T)

Wick's chronological product (WCP) takes into account the commuting or anti-commuting nature of the operators involved. For commuting operators it is the same as the DCP. But, in the case of fermion operators, it differs from DCP by a phase factor:

$$T\{\hat{A}(x)\hat{B}(x')\ldots\} = (-1)^N P\{\hat{A}(x)\hat{B}(x')\ldots\}, \tag{11.172}$$

where N = the number of interchanges of pairs of fermion operators involved in $P\{\ \}$. Thus, for two fermion operators, we have,

$$T\{\hat{A}(x)\hat{B}(x')\} = \begin{cases}\hat{A}(x)\hat{B}(x'), & \text{if } t > t'\\ -\hat{B}(x')\hat{A}(x), & \text{if } t < t'\end{cases} \tag{11.172a}$$

or, using $\epsilon(x_0)$,

$$T\{\hat{A}(x)\hat{B}(x')\} = \epsilon(x_0-x'_0)P\{\hat{A}(x)\hat{B}(x')\} = -T\{\hat{B}(x')\hat{A}(x)\} \tag{11.172b}$$

In general,

$$T\{\hat{A}(x)\hat{B}(x')\} = \pm T\{\hat{B}(x')\,\hat{A}(x)\}, \tag{11.172c}$$

where the upper (lower) sign applies to boson (fermion) operators.

Contraction

Using Eqs. (11.172a, c), (11.170) and (11.169b), we obtain,

$$T\{\hat{A}(x_1)\hat{B}(x_2)\} = :\hat{A}(x_1)\hat{B}(x_2): + \underbrace{\hat{A}(x_1)\hat{B}}(x_2), \tag{11.173}$$

where,
$$\underbrace{\hat{A}(x_1)\hat{B}}(x_2) = \langle\Phi_0 \mid T\{\hat{A}(x_1)\hat{B}(x_2)\} \mid \Phi_0\rangle. \tag{11.174}$$

and is known as the *contraction* of $\hat{A}$ and $\hat{B}$. Obviously, $\underbrace{\hat{A}(x_1)\hat{B}}(x_2)$ will vanish unless one of the operators creates a particle which the other operator annihilates. Therefore, the Feynman diagram corresponding to $\underbrace{\hat{A}(x_1)\hat{B}}(x_2)$ will be an *internal line* [that is, one beginning at a vertex and ending at another vertex (Fig. 11.7)]. It is referred to as a *Feynman propagator* since it represents the creation of a (virtual) particle or antiparticle at one vertex, its propagation to the other vertex and its absorption there. A mathematical expression for the Feynman propagator for the case of each of the three relativistic fields considered in Section 11.4 could be obtained as follows:

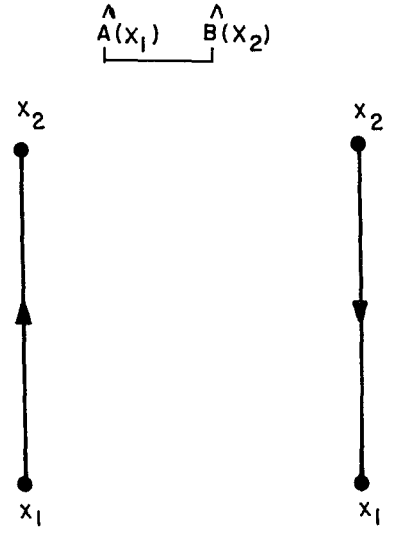


Fig. 11.7. Feynman diagram corresponding to $\underbrace{\hat{A}(x_1)\hat{B}}(x_2)$.

From (11.172) and (11.171[2]) we have, in the case of a (Hermitian) Klein-Gordon field,

$$T\{\hat{\Phi}(x)\Phi(x')\} = P\{\hat{\Phi}(x)\Phi(x')\}$$

$$= \frac{1}{2}[1 + \epsilon\ (x_0 - x'_0)]\hat{\Phi}(x)\hat{\Phi}(x') + \frac{1}{2}[1 - \epsilon\ (x_0 - x'_0)]\hat{\Phi}(x')\hat{\Phi}(x). \tag{11.175}$$

Also, using (11.95[1]),

$$\langle \Phi_0 \mid \hat{\Phi}(x)\,\hat{\Phi}(x') \mid \Phi_0 \rangle = \langle \Phi_0 \mid \hat{\Phi}^{(+)}(x)\hat{\Phi}^{(-)}(x') \mid \Phi_0 \rangle$$

$$= \frac{1}{(2\pi)^3}\int d^3\mathbf{k}\,\frac{\exp\,(ik(x-x'))}{2\omega_{\mathbf{k}}}$$

$$= i\Delta^{(+)}(x-x'), \tag{11.176a}$$

$$\langle \Phi_0 \mid \hat{\Phi}(x')\hat{\Phi}(x) \mid \Phi_0 \rangle = i\Delta^{(+)}(x'-x) = -i\Delta^{(-)}(x-x'). \tag{11.176b}$$

Thus,

$$\underbrace{\hat{\Phi}(x)\hat{\Phi}(x')} = \langle \Phi_0 \mid T\{\hat{\Phi}(x)\hat{\Phi}(x')\} \mid \Phi_0 \rangle$$

$$= \frac{1}{2}\Delta_F(x-x'), \tag{11.177}$$

where

$$\Delta_F(x) = [1 + \in (x_0)]\, i\Delta^{(+)}(x) - [1 - \in (x_0)]\, i\Delta^{(-)}(x) \tag{11.178a}$$

$$= \pm 2i\Delta^{(+)}(x),\ x_0 \gtrless 0. \tag{11.178b}$$

In the case of the Dirac field, we have from (11.172b) and (11.171[2]),

$$\underbrace{\hat{\psi}_\alpha(x)\,\hat{\bar{\psi}}_\beta(x')} = \langle \Phi_0 \mid T\{\hat{\psi}_\alpha(x)\,\hat{\bar{\psi}}_\beta(x')\} \mid \Phi_0 \rangle$$

$$= \frac{1}{2}[1 + \in (x_0 - x_0')]\,\langle \Phi_0 \mid \psi_\alpha(x)\,\hat{\bar{\psi}}_\beta(x') \mid \Phi_0 \rangle$$

$$-\frac{1}{2}[1 - \in (x_0 - x_0')]\,\langle \Phi_0 \mid \hat{\bar{\psi}}_\beta(x')\hat{\psi}_\alpha(x) \mid \Phi_0 \rangle. \tag{11.179[1]}$$

And using Eqs. (11.135a, b) we get,

$$\langle \Phi_0 \mid \hat{\psi}_\alpha(x)\,\hat{\bar{\psi}}_\beta(x') \mid \Phi_0 \rangle = \langle \Phi_0 \mid \hat{\psi}_\alpha^{(+)}(x)\,\hat{\bar{\psi}}_\beta^{(-)}(x') \mid \Phi_0 \rangle$$

$$= -iS_{\alpha\beta}^{(+)}(x-x'), \tag{11.180a}$$

$$\langle \Phi_0 \mid \hat{\bar{\psi}}_\beta(x')\hat{\psi}_\alpha(x) \mid \Phi_0 \rangle = \langle \Phi_0 \mid \hat{\bar{\psi}}_\beta^{(+)}(x')\hat{\psi}_\alpha^{(-)}(x) \mid \Phi_0 \rangle$$

$$= -iS_{\alpha\beta}^{(-)}(x-x'), \tag{11.180b}$$

so that,

$$\underbrace{\hat{\psi}_\alpha(x)\,\hat{\bar{\psi}}_\beta(x')} = -\frac{1}{2}S_{F_{\alpha\beta}}(x-x'), \tag{11.179[2]}$$

$$S_{F_{\alpha\beta}} = [1 + \in (x_0)]\, iS_{\alpha\beta}^{(+)}(x) - [1 - \in (x_0)]\, iS_{\alpha\beta}^{(-)}(x)$$

$$= (\gamma_\mu \partial_\mu - m)\Delta_F(x). \tag{11.181}$$

Here, the last step follows from Eqs. (11.136[1]) and (11.178a). Also,

$$\overbrace{\hat{\psi}_\beta(x')\hat{\psi}_\alpha(x)} = -\overbrace{\hat{\psi}_\alpha(x)\hat{\bar{\psi}}_\beta(x')}, \tag{11.182a}$$

$$\overbrace{\hat{\psi}_\alpha(x)\hat{\psi}_\beta(x')} = \overbrace{\hat{\bar{\psi}}_\alpha(x)\hat{\bar{\psi}}_\beta(x')} = 0. \tag{11.182b}$$

The propagator for the electromagnetic field may be derived analogously to that for the Klein-Gordon field. We get,

$$\overbrace{\hat{A}_\mu(x)\hat{A}_\nu(x')} = \langle \Phi_0 \mid P\{\hat{A}_\mu(x)\hat{A}_\nu(x')\} \mid \Phi_0 \rangle$$

$$= \frac{1}{2} D_F(x - x')\delta_{\mu\nu}, \tag{11.183}$$

where, [see Eq. (11.156^2)],

$$D_F(x) = Lt_{m \to 0} \Delta_F(x, m). \tag{11.184}$$

Wick's Theorem

The normal product of a contraction is obviously the contraction itself (since a contraction does not contain creation or annihilation operators):

$$:\overbrace{\hat{A}\hat{B}}: = \overbrace{\hat{A}\hat{B}}. \tag{11.185a}$$

In general,

$$:(\overbrace{\hat{A}\hat{B}\hat{C}\hat{D}\hat{E}\hat{F}\ldots\hat{J}\hat{K}\hat{L}\hat{M}}\ldots): = (-1)^N \overbrace{\hat{A}\hat{K}}\ \overbrace{\hat{B}\hat{M}}\ \overbrace{\hat{D}\hat{F}} : (\hat{C}\hat{E}\ldots\hat{J}\hat{L}\ldots): \tag{11.185b}$$

That is, the normal product of a product of operators some of which are contracted is the product of the contractions and the normal product of the uncontracted operators. Expression (11.185b) is called a *generalized normal product.*

The concept of the generalized normal product can be utilised to extend Eq. (11.173) to the case of an arbitrary number of operators. The result is,

$$\begin{aligned} &T\{\hat{A}\hat{B}\hat{C}\hat{D}\ldots\hat{W}\hat{X}\hat{Y}\hat{Z}\} \\ &= :(\hat{A}\hat{B}\hat{C}\hat{D}\ldots\hat{W}\hat{X}\hat{Y}\hat{Z}): \\ &\quad + :(\overbrace{\hat{A}\hat{B}}\hat{C}\hat{D}\ldots\hat{W}\hat{X}\hat{Y}\hat{Z}): + :(\overbrace{\hat{A}\hat{B}\hat{C}}\hat{D}\ldots): + \ldots + :(\hat{A}\hat{B}\ldots\hat{W}\hat{X}\overbrace{\hat{Y}\hat{Z}}): \\ &\quad + :(\overbrace{\hat{A}\hat{B}}\overbrace{\hat{C}\hat{D}}\ldots\hat{W}\hat{X}\hat{Y}\hat{Z}): + \ldots + :(\hat{A}\hat{B}\hat{C}\hat{D}\ldots\overbrace{\hat{W}\hat{X}}\overbrace{\hat{Y}\hat{Z}}): \\ &\quad + \ldots\ldots\ldots\ldots\ldots \\ &\quad + (\overbrace{\hat{A}\hat{B}}\overbrace{\hat{C}\hat{D}}\ldots\overbrace{\hat{W}\hat{X}}\overbrace{\hat{Y}\hat{Z}}): \end{aligned} \tag{11.186}$$

This equation gives the WCP of a product of operators in terms of the generalized normal products. It is known as the *Wick's theorem*[21]. It could be proved by induction starting from Eq. (11.173).

21. Wick, G.C. *Phys. Rev.* **80**, 268 (1950).

Electromagnetic Coupling

The interaction between a Dirac field and an (external) electromagnetic field can be described in terms of the modification of the 4-momentum of a Dirac particle arising from the coupling of the two fields. This modification is given by

$$\hat{p}_\mu \to p_\mu' = \hat{p}_\mu - e\hat{A}_\mu, \text{ or, } \partial_\mu \to \partial_\mu - ie\hat{A}_\mu. \tag{11.189}$$

Making this substitution in (11.103a), we get for the Lagrangian density of the system the expression,

$$\hat{\mathcal{L}}(x) = \hat{\mathcal{L}}_D(x) + \hat{\mathcal{L}}_I(x), \tag{11.190}$$

with $\hat{\mathcal{L}}_D$ given by Eq. (11.103a), and

$$\hat{\mathcal{L}}_I(x) = ie\,\hat{\bar{\psi}}(x)\gamma_\mu\hat{\psi}(x)\hat{A}_\mu(x) = e\hat{j}_\mu\hat{A}_\mu. \tag{11.191}$$

(We see that the coupling constant in this case is e). The corresponding Hamiltonian density is given by

$$\mathcal{H}(x) = \hat{\pi}(x)\,\dot{\hat{\psi}}(x) - \hat{\mathcal{L}}(x) = \hat{\mathcal{H}}_{D_{(x)}} + \hat{\mathcal{H}}_I(x), \tag{11.192}$$

where $\hat{\mathcal{H}}_D(x)$ is given by Eq. (11.120) and

$$\hat{\mathcal{H}}_I(x) = -\hat{\mathcal{L}}_I(x) = -ie\,\hat{\bar{\psi}}(x)\gamma_\mu\hat{A}_\mu(x)\hat{\psi}(x) \tag{11.193a}$$

$$= -ie(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_x$$

$$= -ie{:}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_x{:} \tag{11.193b}$$

Here,

$$\hat{\mathbf{A}} = \gamma_\mu\hat{A}_\mu. \tag{11.194}$$

(11.193b) follows from (11.193a) by the application of (11.170) and of the fact that the vacuum expectation value of $(\hat{\bar{\psi}}\mathbf{A}\hat{\psi})$ is zero.
The Hamiltonian of the system is,

$$\hat{H} = \hat{H}_D + \hat{H}_I, \tag{11.195}$$

where

$$\hat{H}_I(t) = \int d^3x\,\hat{\mathcal{H}}_I(x). \tag{11.196}$$

The Scattering Matrix

The basic equation to be solved in the case of a coupled system is,

$$i\partial_t\Phi(t) = \hat{H}_I(t)\Phi(t), \tag{11.197}$$

where both the state vector Φ and the interaction Hamiltonian $\hat{H}_I$ are assumed to be in the interaction picture[22] [see Eqs. (4.40) and (8.157) and Problem 4.4].

22. As explained in Section 4.1C, the interaction picture is advantageous when the Hamiltonian can be split up into two parts as in (11.195). Also, the interaction picture is better suited, as compared with the Schrödinger picture, for a covariant formulation because of the time-dependence of *both* the operators and the state vectors in the former.

As described in Section 8.4, $\Phi(t)$ can be expressed as an infinite (perturbation) series:

$$\Phi(t) = \hat{U}_I(t, t_0)\Phi(t_0), \tag{11.198^1}$$

where [see Eqs. (8.159^3) and (8.160a) and Problem 8.12],

$$\hat{U}_I(t, t_0) = \hat{1} + \sum_{n=1}^{\infty} \hat{U}_I^{(n)}(t, t_0), \tag{11.199}$$

$$\hat{U}_I^{(n)}(t, t_0) = (-i)^n \int_{t_0}^{t} dt_1 \int_{t_0}^{t} dt_2 \dots \int_{t_0}^{t_{n-1}} dt_n \{\hat{H}_I(t_1) \dots \hat{H}_I(t_n)\}$$

$$= \frac{(-i)^n}{n!} \int_{t_0}^{t} dt_1 \int_{t_0}^{t} dt_2 \dots \int_{t_0}^{t} dt_n P\{\hat{H}_I(t_1) \dots \hat{H}_I(t_n)\} \tag{11.199a}$$

In a scattering problem the initial state $\Phi(t_0)$ corresponds to $t_0 = -\infty$ (long before scattering takes places) and the final state $\Phi(t)$ to $t = +\infty$ (long after scattering has occurred). Thus, Eq. (11.198^1) becomes in this case,

$$\Phi(\infty) = S\Phi(-\infty), \tag{11.198^2}$$

where

$$S = U_I(+\infty, -\infty), \tag{11.200^1}$$

and is called the *scattering (or S–)* matrix [$U_I(\infty, -\infty)$ is the matrix representation of $\hat{U}_I(\infty, -\infty) \equiv \hat{S}$]. In (11.198^2). $\Phi(-\infty)$ stands for a set of initial states and $\Phi(\infty)$ for a set of final states:

$$\Phi(-\infty) = \begin{pmatrix} \Phi_1(-\infty) \\ \Phi_2(-\infty) \\ \vdots \\ \Phi_i(-\infty) \\ \vdots \end{pmatrix}; \quad \Phi(\infty) = \begin{pmatrix} \Phi_1(-\infty) \\ \Phi_2(-\infty) \\ \vdots \\ \Phi_j(-\infty) \\ \vdots \end{pmatrix},$$

so that the ji-th matrix element of S:

$$S_{ji} = \langle \Phi_j(\infty) \mid \hat{U}_I(\infty, -\infty) \mid \Phi_i(-\infty) \rangle, \tag{11.201}$$

gives the probability amplitude for the transition $\Phi_i(-\infty)$ to $\Phi_j(\infty)$. A complete knowledge of S, therefore, enables us to predict the probability for transition from any initial state to any final state. Finding S is, thus, equivalent to solving Eq. (11.197). This is the basis of the S-matrix formalism.[23]

In Eqs. (11.198^2) and (11.201) the initial and the final states are identified with the eigenstates of the free Hamiltonian ($\hat{H}_D$ in (11.195)) and, thus, with the 'bare' particles. However, in a scattering experiment the initial and final states are those of the physical particles which result from the interaction of the bare particles with their own fields or photon clouds. The inconsistency is resolved by assuming an adiabatic switching on and off of the coupling constant. That is, the coupling

23. The S-matrix formalism was originally developed by Wheeler [Wheeler, J.A. *Phys. Rev.* **52**, 1107 (1937)], but its further development and application to the interaction of elementary particles are due to Heisenberg [Heisenberg, W. *Z. Physik*, **120**, 513, 673 (1943)].

constant is zero at $t = -\infty$, increases adiabatically, reaching its full value at $t = -T$, retains this full value during $-T \le t \le T$ and thereafter decreases to the value zero at $t = \infty$. The scattering takes place during the time $-\tau < t < \tau$, where $\tau \ll T$.

Now, from (11.200[1]), (11.199) and (11.199a), we have,

$$\hat{S} = \sum_{n=0}^{\infty} \hat{S}^{(n)}, \tag{11.200[2]}$$

where $$\hat{S}^{(n)} = \frac{(-i)^n}{n!} \int_{-\infty}^{+\infty} dt_1 \ldots \int_{-\infty}^{+\infty} dt_n P\{\hat{H}_I(t_1) \ldots \hat{H}_I(t_n)\}. \tag{11.202[1]}$$

$$= \frac{(-i)^n}{n!} \int d^4x_1 \ldots \int d^4x_n P\{\hat{\mathcal{H}}_I(x_1) \ldots \hat{\mathcal{H}}_I(x_n)\}. \tag{11.203}$$

where, $d^4x_j = d^3\mathbf{x}_j dt_j$.

Since, according to (11.193), $\hat{\mathcal{H}}_I(x)$ is bilinear in the fermion operators,

$$P\{\hat{\mathcal{H}}_I(x_1) \ldots \hat{\mathcal{H}}_I(x_n)\} = T\{\hat{\mathcal{H}}_I(x_1) \ldots \hat{\mathcal{H}}_I(x_n)\}$$

$$= (-ie)^n T\left\{:(\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_1}: \ldots :(\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_n}:\right\} \tag{11.204}$$

The Wick's chronological product in (11.204) can be expanded in terms of the generalized normal products with the help of Wick's theorem [Eq. (11.186)]. Since the contraction of a normal product is zero, we can write,

$$T\{:(\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_1}: \ldots :\left\{(\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_n}:\right\} = T'\left\{(\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_2} \ldots (\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_n}\right\} \tag{11.205}$$

where T' contains no equal time contractions. Substituting from (11.204) and (11.205) in (11.202[1]), we get,

$$\hat{S}^{(n)} = \frac{(-1)^n e^n}{n!} \int d^4x_1 \ldots \int d^4x_n T'\left\{(\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_1} \ldots (\hat{\bar{\psi}}\hat{A}\hat{\psi})_{x_n}\right\} \tag{11.202[1]}$$

We will consider the scattering processes corresponding to the lowest few values of n:

n = 0:

$$\hat{S}^{(0)} = \hat{1}. \tag{11.206}$$

This represents transitions from a state to the same state (no scattering). The Feynman diagram contains no vertex (Fig. 11.8).

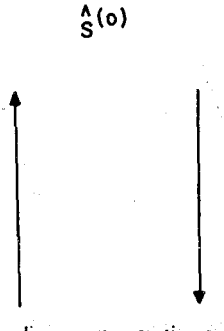


Fig. 11.8. Feynman diagram representing zero-order processes.

n = 1:

$$\hat{S}^{(1)} = -e \int d^4x T'\left\{\left(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi}\right)_x\right\} = \hat{S}_a^{(1)} + \hat{S}_b^{(1)} + \hat{S}_c^{(1)} + \hat{S}_d^{(1)} \tag{11.207}$$

where

$$\hat{S}_a^{(1)} = -e \int d^4x \left(\hat{\bar{\psi}}^{(+)}\hat{\mathbf{A}}\hat{\psi}^{(+)}\right)_x, \tag{11.207a}$$

$$\hat{S}_b^{(1)} = -e \int d^4x \left(\hat{\bar{\psi}}^{(+)}\hat{\mathbf{A}}\hat{\psi}^{(-)}\right)_x, \tag{11.207b}$$

$$\hat{S}_c^{(1)} = -e \int d^4x \left(\hat{\bar{\psi}}^{(-)}\hat{\mathbf{A}}\hat{\psi}^{(+)}\right)_x, \tag{11.207c}$$

$$\hat{S}_d^{(1)} = -e \int d^4x \left(\hat{\bar{\psi}}^{(-)}\hat{\mathbf{A}}\hat{\psi}^{(-)}\right)_x, \tag{11.207d}$$

We recall [Eq. (11.118)] that,

$\hat{\bar{\psi}}^{(+)}$ is linear in positron annihilation operator,

$\hat{\bar{\psi}}^{(-)}$ is linear in electron creation operator,

$\hat{\psi}^{(+)}$ is linear in electron annihilation operator,

$\hat{\psi}^{(-)}$ is linear in positron creation operator.

The Feynman graphs corresponding to these are shown in Fig. 11.9.

Thus, $\hat{S}_a^{(1)}$ can cause transition from a state Φ_i to a state Φ_f which differs from Φ_i by the absence of an electron-positron pair. That is, $\hat{S}_a^{(1)}$ corresponds to pair annihilation (Fig. 11.4(b)). Similarly, $\hat{S}_b^{(1)}$, $\hat{S}_c^{(1)}$ and $\hat{S}_d^{(1)}$ correspond respectively to positron scattering (Fig. 11.3(b)), electron scattering (Fig. 11.3(a)) and pair creation (Fig. 11.4(a)). Conservation of energy and momentum, however, prevents the above processes from taking place. Therefore, the lowest order term in the S-matrix expansion that corresponds to physical processes is of second order.

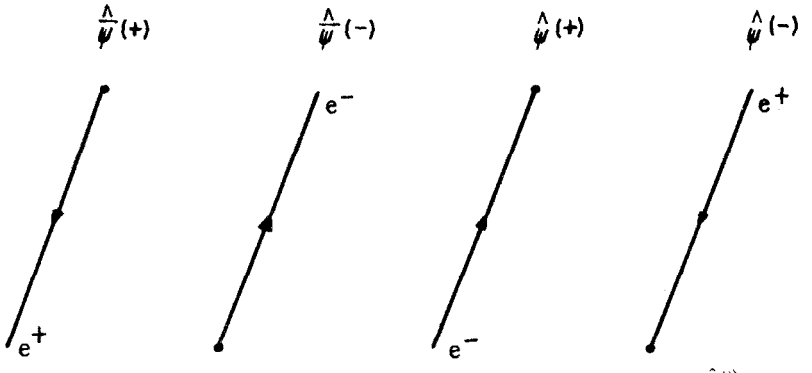


Fig. 11.9 Feynman diagrams corresponding to the field operators $\hat{\psi}^{(\pm)}$ and $\hat{\bar{\psi}}^{(\pm)}$.

n = 2:

$$\hat{S}^{(2)} = \frac{e^2}{2}\int d^4x_1 \int d^4x_2 T^*\left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\}$$

$$= \sum_{i=A}^{F} \hat{S}_i^{(2)}, \tag{11.208}$$

where,

$$\hat{S}_A^{(2)} = \frac{e^2}{2}\int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} : \tag{11.208a}$$

$$\hat{S}_B^{(2)} = e^2\int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} : \tag{11.208b}$$

$$\hat{S}_C^{(2)} = \frac{e^2}{2}\int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} : \tag{11.208c}$$

$$\hat{S}_D^{(2)} = e^2\int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} : \tag{11.208d}$$

$$\hat{S}_E^{(2)} = \frac{e^2}{2}\int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} : \tag{11.208e}$$

$$\hat{S}_F^{(2)} = \frac{e^2}{2}\int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} : \tag{11.208f}$$

Eq. (11.208a-f) follow by the application of the Wick's theorem to the WCP in (11.208) and by noting that

$$\int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}\right\} :$$

$$= \int\int d^4x_1 d^4x_2 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} :$$

$$= \int\int d^4x_2 d^4x_1 : \left\{(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_1}(\hat{\bar{\psi}}\hat{\mathbf{A}}\hat{\psi})_{x_2}\right\} : \tag{11.209}$$

$\hat{S}_A^{(2)}$ does not contain a propagator between the vertices x_1 and x_2 and thus, consists of two *unconnected* basic vertex parts. Therefore it corresponds to the same type of processes as those represented by $\hat{S}^{(1)}$.

Expanding the normal product in $\hat{S}_B^{(2)}$ in terms of the positive and the negative frequency parts of $\hat{\psi}$, $\hat{\bar{\psi}}$ and $\hat{\mathbf{A}}$ we have,

$$\hat{S}_B^{(2)} = \sum_{i=1}^{8} \hat{S}_{Bi}^{(2)},$$

with (here, for the sake of ease of interpretation, we keep together the operators referring to the same vertex),

$$\hat{S}^{(2)}_{B1} = e^2 \int\int d^4x_1 d^4x_2 \hat{\bar{\psi}}^{(-)}(x_2)\hat{A}^{(-)}(x_2)\underbrace{\hat{\psi}(x_2)\hat{\bar{\psi}}}(x_1)\hat{A}^{(+)}(x_1)\hat{\psi}^{(+)}(x_1),$$

$$\hat{S}^{(2)}_{B2} = e^2 \int\int d^4x_1 d^4x_2 \hat{\bar{\psi}}^{(-)}(x_2)\hat{A}^{(+)}(x_2)\underbrace{\hat{\psi}(x_2)\hat{\bar{\psi}}}(x_1)\hat{A}^{(-)}(x_1)\hat{\psi}^{(+)}(x_1),$$

$$\hat{S}^{(2)}_{B3} = e^2 \int\int d^4x_1 d^4x_2 \hat{\psi}^{(-)}(x_2)\hat{A}^{(-)}(x_2)\underbrace{\hat{\psi}(x_1)\hat{\bar{\psi}}}(x_2)\hat{A}^{(+)}(x_1)\hat{\bar{\psi}}^{(+)}(x_1),$$

$$\hat{S}^{(2)}_{B4} = e^2 \int\int d^4x_1 d^4x_2 \hat{\psi}^{(-)}(x_2)\hat{A}^{(+)}(x_2)\underbrace{\psi(x_1)\hat{\bar{\psi}}}(x_2)\hat{A}^{(-)}(x_1)\hat{\bar{\psi}}^{(+)}(x_1),$$

$$\hat{S}^{(2)}_{B5} = e^2 \int\int d^4x_1 d^4x_2 \hat{\bar{\psi}}^{(-)}(x_2)\hat{A}^{(+)}(x_2)\underbrace{\psi(x_2)\hat{\bar{\psi}}}(x_1)\hat{A}^{(+)}(x_1)\hat{\psi}^{(-)}(x_1),$$

$$\hat{S}^{(2)}_{B6} = e^2 \int\int d^4x_1 d^4x_2 \hat{\bar{\psi}}^{(+)}(x_2)\hat{A}^{(-)}(x_2)\underbrace{\hat{\psi}(x_2)\hat{\bar{\psi}}}(x_1)\hat{A}^{(-)}(x_1)\hat{\psi}^{(+)}(x_1),$$

$$\hat{S}^{(2)}_{B7} = e^2 \int\int d^4x_1 d^4x_2 \hat{\bar{\psi}}^{(-)}(x_2)\hat{A}^{(-)}(x_2)\underbrace{\hat{\psi}(x_2)\hat{\bar{\psi}}}(x_1)\hat{A}^{(+)}(x_1)\hat{\psi}^{(-)}(x_1),$$

$$\hat{S}^{(2)}_{B8} = e^2 \int\int d^4x_1 d^4x_2 \hat{\bar{\psi}}^{(-)}(x_2)\hat{A}^{(-)}(x_2)\underbrace{\hat{\psi}(x_2)\hat{\bar{\psi}}}(x_1)\hat{A}^{(-)}(x_1)\hat{\psi}^{(-)}(x_1).$$

Now, a contraction corresponds to an internal line in a Feynman diagram while the uncontracted operators are represented by external lines. Thus, $\hat{S}^{(2)}_B$ represents processes that correspond to Feynman diagrams with two external fermion lines, two external photon lines and one internal fermion line. Also, we see from (11.179[1]) and (11.180a, b) that $\underbrace{\hat{\psi}(x_2)\hat{\bar{\psi}}}(x_1)$ is the propagator for a virtual electron from x_1 to x_2 if $t_1 < t_2$ and for a virtual positron from x_2 to x_1 if $t_2 < t_1$. Referring then to Fig 11.9, it is easy to see that the Feynman diagrams corresponding to $\hat{S}^{(2)}_{B1}$ and $\hat{S}^{(2)}_{B2}$ are as in Fig. (11.5). That is, $\hat{S}^{(2)}_{B1}$ and $\hat{S}^{(2)}_{B2}$ represent Compton scattering by electrons. Similarly, $\hat{S}^{(2)}_{B3}$ and $\hat{S}^{(2)}_{B4}$ correspond to Compton scattering by positrons (Fig. 11.10), $\hat{S}^{(2)}_{B5}$ and $\hat{S}^{(2)}_{B6}$ respectively to two-quantum pair creation (Fig. 11.6(a)) and annihilation (Fig. 11.6(b)). $\hat{S}^{(2)}_{B7}$ and $\hat{S}^{(2)}_{B8}$ do not correspond to any physical processes as they would lead to violation of the conservation laws.

A similar analysis of the other terms in (11.208) could be made. We give below the main conclusions:

$\hat{S}^{(2)}_C$ corresponds to processes represented by Feynman diagrams consisting of four external fermion lines and one internal photon line. Two of these processes are depicted in Fig. 11.11. (a) represents electron-electron scattering (Moller scattering) while (b) represents pair annihilation in the field of a positron. The other processes corresponding to $\hat{S}^{(2)}_C$ are electron-positron scattering (Bhabha scattering), positron-positron scattering, pair creation in the field of a positron and pair creation and annihilation in the field of an electron.

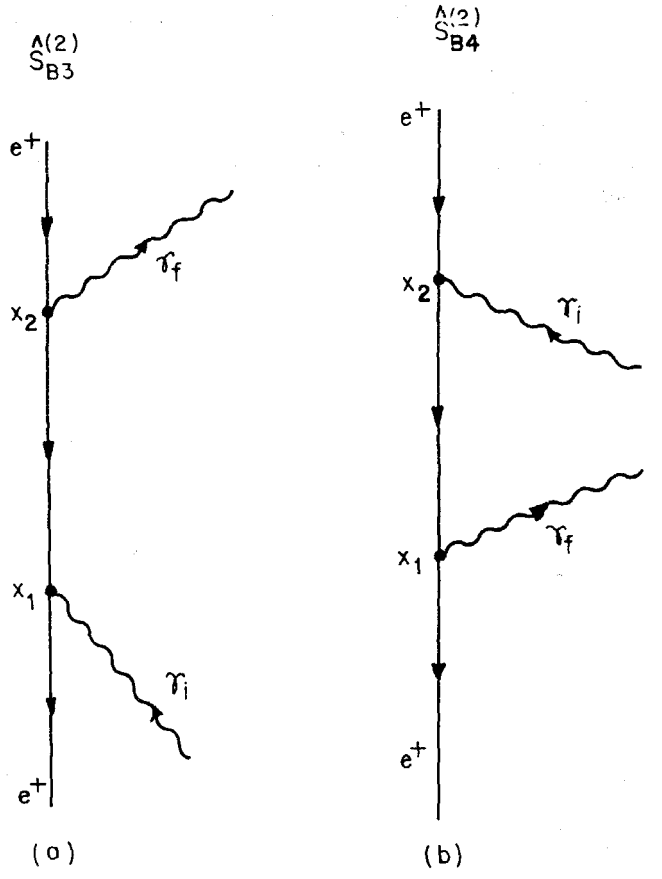


Fig. 11.10. Compton scattering by positrons.

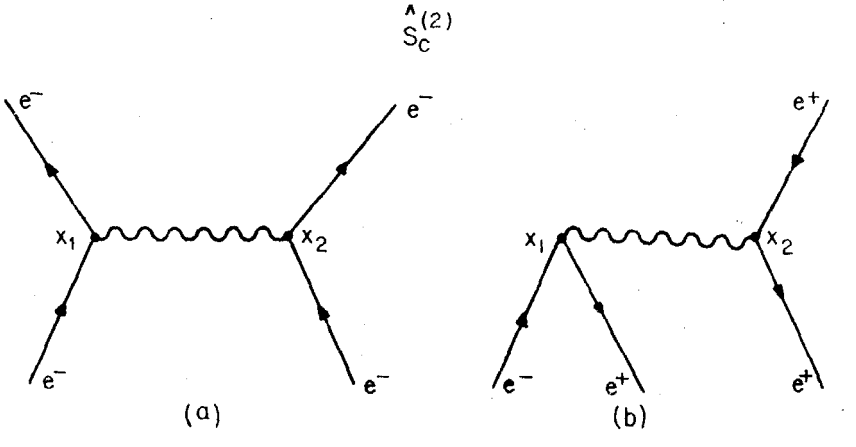


Fig. 11.11. Processes represented by $\hat{S}^{(2)}_C$ [Eq. (11.208c)]: (a) electron-electron scattering, (b) pair-annihilation in the field of a positron.

Processes corresponding to $\hat{S}_D^{(2)}$ have two external fermion lines connected by an internal fermion line and an internal photon line as shown in Fig. 11.12. Fig. 11.12(a) represents the interaction of an electron with the photon field. This is referred to as the *electron self-energy diagram* as it represents a process by which a *bare* electron converts itself into a *physical* electron. (The mass of the physical electron is larger than that of the bare electron.) Similarly Fig. 11.12(b) is the positron self-energy diagram.

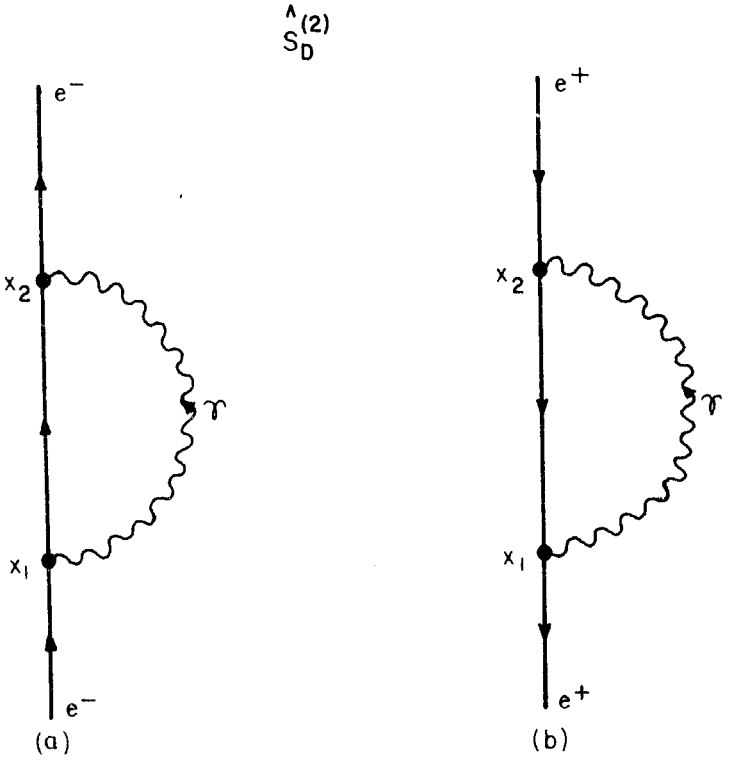


Fig. 11.12. Processes corresponding to $\hat{S}_D^{(2)}$ [Eq. (11.208d)]: (a) electron self-energy, (b) positron self-energy.

$\hat{S}_E^{(2)}$ corresponds to *photon self energy* (Fig. 11.13). As seen from the figure, the photon creates a virtual electron-positron pair which later annihilate giving back the photon. The vacuum will contain such virtual electron-positron pairs. An external electromagnetic field can affect the distribution of these virtual electron-positron pairs. The effect is referred to as *vacuum polarization.*

Fig. 11.14 gives the Feynman diagram corresponding to $\hat{S}_F^{(2)}$. Since there are no external lines, it does not lead to any observable phenomenon.

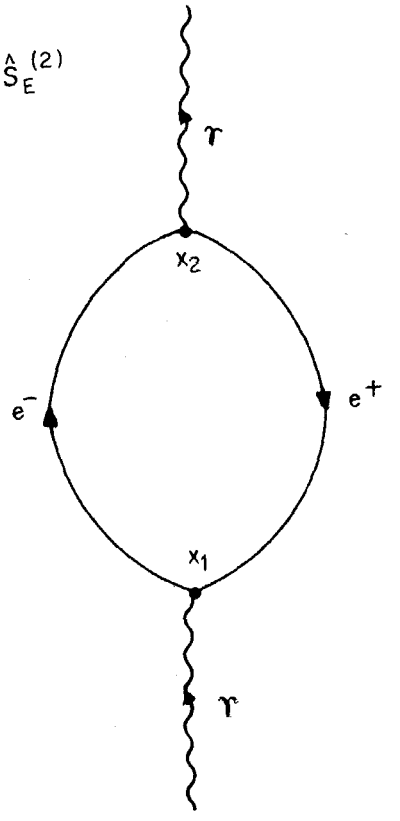


Fig. 11.13. Process represented by $S_E^{(2)}$ [Eq. (11.208e)]: Photon self-energy (vacuum polarization).

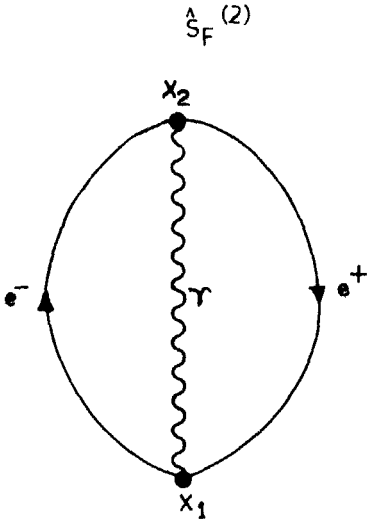


Fig. 11.14. Diagrammatic representation of $S_F^{(2)}$ [Eq. (11.208f)].

The above discussion should suffice to give some familiarity with the technique of Feynman diagrams and their usefulness in the treatment of problems involving interaction between fields.

REFERENCES

1. Mandl, F. *Introduction to Quantum Field Theory* (Interscience, New York 1959), Chapters 1-14.
2. Lurie, D. *Particles and Fields* (Interscience, New York 1968).
3. Muirhead, H. *The Physics of Elementary Particles* (Pergamon, Oxford 1968), Chapters 4, 7 and 8.
4. Schiff, L.I. *Quantum Mechanics* (McGraw-Hill—Kogakusha, Tokyo 1968), Chapter 14.
5. Roman, P. *Advanced Quantum Theory* (Addison-Wesley, Massachusetts 1962), Section 1.7.

CHAPTER 12

THE INTERPRETATIONAL PROBLEM

In Chapter 1, we had briefly referred to the problem associated with the physical interpretation of the quantum formalism. As Heisenberg put it[1]: "the mathematical equipment of the theory was complete in its most important parts by the middle of 1926, but the physical significance was still extremely unclear". It was in recognition of this fact that a major part of the Fifth Solvay Conference[2] held at Brussels in 1927 was devoted to a debate on the problem of the interpretation of quantum mechanics. The debate between Albert Einstein and Niels Bohr started at the Conference and continued both within and without the later conferences, resulted in the emergence of the two opposing schools of interpretation - the *Copenhagen* and the *Statistical Ensemble* - sketched in Chapter 1. In this chapter, we propose to give a fuller account of the two schools of interpretation. We will also give an account of the Hidden-Variable Theories and the associated developments which arose as an off-shoot of the interpretational problem.

12.1 THE EINSTEIN-PODOLSKY-ROSEN (EPR) PARADOX

As illustrated by the double-slit interference experiment on electrons described in Chapter 1, the main point of dispute in the interpretation of quantum phenomena is concerned with the nature of the underlying physical reality: Is the electron a particle, a wave, both or neither? A significant aspect of the experimental result (interference phenomenon) is that this dilemma regarding the nature of electrons (or other such physical entities) intrudes itself only when we think in terms of the (classical) trajectories of the individual electrons involved in the process. If we are content with a statistical description in terms of ensembles - the interference pattern as a statistical distribution of electrons - the concept of electrons as being particles traversing individual, but *unpredictable*, trajectories is alright. The fundamental problem of the interpretation of quantum phenomena, thus, could also be phrased as follows: "Do the laws of quantum mechanics provide a

---

1. Heisenberg W. in *Neils Bohr and the Development of Physics*, (ed) W. Paul (Pergamon, Oxford 1955), p. 13.
2. See Mehra, J. *The Solvay Conferences on Physics* (D.Reidel Publishing Co, Boston 1975), Chapter 6.

complete description of an individual system, or do they embody only the statistical laws governing ensembles?" This was forcefully brought out in a paper entitled "Can quantum-mechanical description of physical reality be considered complete?" By Einstein, Podolsky and Rosen[3] published in 1935. This paper is, in fact, a culmination of the debate between Neils Bohr and Einstein started at the fifth Solvay Conference. The paper begins with the definitions of *Completeness* and of *Physical Reality:*

(EPR.1) A *necessary* condition for the completeness of a theory is that "every element of physical reality must have a counterpart in the theory".

(EPR.2) A *sufficient* condition to identify an element of physical reality is: "If, without in any way disturbing a system, we can predict with certainty (i.e. with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this *physical quantity*".

Then follows the description of a 'thought-experiment'. We give below a brief account of this EPR experiment in its Bohm version[4].

Consider a system consisting of two spin 1/2 particles in a singlet (S = 0) state. The wave function of the system is given by [cf. Eq. (9.15)]

$$\psi_0(1,2) = \frac{1}{\sqrt{2}}\left[x^{(1)}_{+\frac{1}{2}}(\mathbf{a})\, x^{(2)}_{-\frac{1}{2}}(\mathbf{a}) - x^{(1)}_{-\frac{1}{2}}(\mathbf{a}) x^{(2)}_{+\frac{1}{2}}(\mathbf{a})\right], \tag{12.1}$$

where $x^{(i)}_m(\mathbf{a})$ denotes the spin wave function of the ith particle with spin component equal to $m\hbar$ along the direction of the unit vector **a**. Let the system decay into the component particles by angular momentum-conserving interactions. The two particles will fly apart in different directions and the interaction between them will vanish. A measurement of $S^{(1)}_a$ (the component along **a** of the spin $\mathbf{S}^{(1)}$ of particle number 1) will yield the value $\pm(\hbar/2)$. Correspondingly, the second particle will have spin equal to $\mp(\hbar/2)$ along **a**. Thus, we can predict with certainty the value of $S^{(2)}_a$ without in any way disturbing the second particle which is spatially separated from the first particle on which measurement is being made. Therefore, according to (EPR.2) above, the component of $S^{(2)}$ along **a** is an element of physical reality that exists separately for the second particle alone. If so, $S^{(2)}_a$ *must have the same value even before the measurement.* But **a** can be in any arbitrary direction we choose. Consequently, elements of reality must exist for particle number 2 relating to its spin component in any arbitrary direction. In particular, definite values must exist for the z-and the x-components of the spin of particle number 2. This is, however, not permitted by quantum mechanics as $S^{(2)}_x$ and $S^{(2)}_z$

3. Einstein A., Podolsky B. and Rosen N., *Phys. Rev.* **47**, 777 (1935).
4. Bohr, D. *Quantum Theory* (Prentice-Hall, Englewood-Cliff, 1951) p. 614. In the original EPR experiment, the incompatible variables are the position and the momentum, but in the Bohm version they are the components of the spin.

are represented by non-commuting Hermitian operators. [See Eq. (10.35a) and the discussion following Postulate II in Chapter 3]. Therefore, every element of physical reality does not have a counter part in quantum mechanics and, as per criterion (EPR.1) above, the answer to the question that forms the title of the paper, is negative.

The foregoing conclusions of the paper has been enunciated in the form of a theorem[5] :

Theorem: *The following two assertions are not compatible with each other:*

(E1). *The description by means of the ψ-function is* **complete.**

(E2). *The real states of spatially separated non-interacting objects are* **independent.**

(E2) is usually referred to as the *Einstein locality postulate.* The EPR experiment, thus, establishes that quantum mechanics is incompatible with the concept of *local realism* which is a basic tenet of classical physics. In other words, we have to regard quantum mechanics either as an *incomplete* theory or as a *non-local* theory with 'spooky actions at a distance'. The latter alternative arises if one adopts the viewpoint that a particular component of $S^{(2)}$ acquires a definite value (and becomes an element of physical reality) as a result of the measurement of the corresponding component of $S^{(1)}$, which requires passing information instantaneously from particle number 1 to particle number 2 though the two are too far apart for direct interaction[6].

Whether we adopt one or the other of the above alternatives, we cannot avoid the *paradox* of arriving at a conclusion prohibited by quantum mechanics (that two non-commuting observables are simultaneously elements of physical reality in a quantum state) starting from premises prescribed by quantum mechanics (namely, representing a state by a wave function) through arguments permitted by quantum mechanics.

The EPR paper has become a classic in the annals of the epistemological and the philosophical foundations of quantum mechanics. This is not only because of its lucid exposition (for the first time) of the basic problem to be tackled in any attempt at a physical interpretation of the quantum formalism but also because of the later developments like the formulation of hidden-variable theories and the enunciation of Bell's theorem that it gave rise to. In fact, the EPR Paradox continues to be a problem for discussion even after fifty years of its publication[7]. In the next two sections we will present Bohr's and Einstein's interpretations of quantum mechanics. In Section 12.4 we will examine the explanation of the EPR paradox from the various points of view and in Section 12.5 we will give a brief account of the hidden-variable theories and the associated developments.

---

5. Einstein A. in *Albert Einstein: Philosopher Scientist*, (ed) P.A. Schilpp (Harper and Row, New York 1959), p. 665.

6. The non-locality involved here is essentially of the same nature as the one involved in the case of an electron passing through slit number 1 in the double-slit interference experiment 'taking notice' of the closed or open state of slit number 2, though in this case there is something physical (the screen bearing the slits) that connects the two slits.

7. See, for example, Selleri, F., *Science Today*, **21**, No. 12, p. 17 (1987); *Quantum Paradoxes and Physical Reality*, (ed.) Alwyn von der Merwe (Kluwer Academic Publishers, Dordricht 1990), Chapter 5; Mermin, N.D. *Physics Today*, **36**, No. 4, p. 38 (1985).

12.2 THE COPENHAGEN INTERPRETATION

The Copenhagen interpretation or *Copenhagen School* derives its name from the fact that the originator and the chief interpreter of this school was Niels Bohr whose headquarters was in Copenhagen. Vis-a-vis Einstein's Ensemble Interpretation to be described in the next section, it could, more appropriately, be termed the *Individual System Interpretation.* Eventhough a variety of viewpoints are attributed to this school, we will discuss only the version which can be traced directly to Niels Bohr[8].

Two of the basic principles that underlie the Copenhagen Interpretation are Heisenberg's *Uncertainty Principle* and Bohr's *Principle of Complementarity.* Therefore we will first give a brief description of these before embarking on Bohr's interpretation of quantum theory.

The Uncertainty Principle

In Section 3.2 we have already given the mathematical derivation of the Uncertainty relationship. However, more than the mathematics, it was the philosophical outlook that accompanied the formulation of the uncertainty principle that influenced the interpretation of quantum mechanics. In fact, the uncertainty principle was the culmination of the realization on the part of Heisenberg that the formalism of quantum mechanics did not admit of ordinary space-time descriptions or causal connection of physical phenomena. Heisenberg wrote to Pauli in 1926[9] (prior to the formulation of the uncertainty principle in the subsequent year): "It makes no sense to speak of a monochromatic wave at a definite instant, or the place of a particle with a definite velocity". The same sentiments were expressed by Dirac and Jordan, the authors of the transformation theory (on which Heisenberg based his derivation of the uncertainty relationship): "One cannot answer any question on the quantum theory, which refers to numerical values for both the q and the p. One would expect, however, to be able to answer questions in which only the q or only the p are given numerical values"[10], while Jordan[11] concluded: "for a given value of q all values of p are equally probable".

Thus the Uncertainty Principle profoundly affected the classical notions of position, velocity and orbit of a particle. The extent to which these were affected are exemplified by the following viewpoints attributed to Heisenberg himself[12] :

---

8. Bohr, N. 'Discussions with Einstein on Epistemological Problems in Atomic Physics', in *Albert Einstein : Philosopher-Scientist* (Tudor Publishing Co. 1949), reproduced in Mehra, J. *The Solvay Conferences on Physics* (D.Reidel Publishing Co. 1975), p. 153; Max Jammer, *The Conceptual Development of Quantum Mechanics* (McGraw-Hill, New York 1966), Ch. 7; Stapp, H.P. *Amer J.Phys.* **40**, 1098 (1972).
9. Max Jammer (Footnote 8) p. 325
10. Dirac, P.A.M. *Proe. Roy. Soc.* (London), **A113**, 621 (1926)
11. Jordan, P. *Zeit. f. Physik,* **40**, 809 (1927)
12. Max Jammer (Footnote 8) pp. 328-330.

> "If one wants to clarify what is meant by 'position of an object', for example an electron, one has to describe an experiment by which the 'position of an electron' can be measured; otherwise this term has no meaning at all ..." "the path comes into existence only when we observe it..." in the strong formulation of the causal law, 'If we know exactly the present, we can predict the future', it is not the conclusion but rather the premise which is false. We cannot know, as a matter of principle, the present in all its detail".

It is obvious that the Uncertainty Principle bestowed on measurement a role in quantum mechanics which it did not have in classical mechanics. Whereas in classical mechanics measurements merely revealed what existed before the measurement was made, in quantum mechanics "observations not only disturb what has to be measured, they produce it"[13].

The Principle of Complementarity

Bohr presented his ideas on complementarity in relation to quantum theory for the first time in a lecture on "The Quantum Postulate and the Recent Development of Atomic Theory" delivered at the International Congress of Physics held at the Instituto Carducci in Como in September 1927. Later he published the paper in *Nature*[14] . The Principle of Complementarity marks the end of Bohr's opposition to Einstein's concept of the light quantum and his (Bohr's) final acceptance of the wave-particle duality. In fact, Bohr's ideas on complementarity seem to have evolved as a result of his attempts at harmonizing the mutually exclusive notions of waves and particles that the dual behaviour of matter and radiation demanded. Bohr was convinced that this harmonization could not be achieved merely by modifying or reinterpreting traditional (classical) concepts. What was needed was a new logical outlook. This he called "complementarity".

Bohr did not attempt an explicit definition of complementarity. Instead, he described the basic ideas underlying it. These could be summarised as follows:

Quantum phenomena are governed by the quantum postulate according to which, to any atomic process an essential individuality, symbolized by Planck's quantum of action, has to be attributed. This means that, in the case of a quantum mechanical (atomic) system, the interaction between the object of observation and the agency of observation (the 'measuring instrument'), and the resulting disturbance of the former, cannot be neglected. As a result, it is impossible to separate the behaviour of the atomic system from the effect of the measuring instrument whose behaviour must be described classically. [Bohr refers to this situation as the indeterminateness of concept of observation, 'in analogy with a similar situation in William James' analysis of the notion of observation in psychology[15] (see Ref.8, p. 349)]. By combining an atomic system with different (measuring)

---

13. This remark is attributed to Jordan [See, Max Jammer, *The Philosophy of Quantum Mechanics* (John Wiley, New York 1974), p. 151].
14. Bohr, N. *Nature*, **121**, 580 (1928)
15. See, Max Jammer (Footnote 8) p. 349

devices, one may observe different aspects of the system. These different aspects could be considered as providing a description of the atomic system in terms of complementary classical pictures. Thus, in a particular experiment, wave nature of radiation may be revealed, whereas in another experimental situation particle aspect of the radiation may be observed. Though (classically) the wave and the particle pictures are mutually exclusive, these are indispensable for a complete (or, *exhaustive*) description of radiation and should, therefore, be considered as complementary to each other.

The uncertainty principle ensures that complementarity does never lead to any contradiction in spite of the logically contradictory nature of the notions involved. For, the uncertainty principle shows that the sharp exhibition of one of such complementary aspects necessitates an experimental set up which is totally different from that required for the exhibition of the other aspect. As a result, there cannot arise any physical situation which reveals simultaneously and sharply both complementary aspects of a phenomenon. Thus, in the double-slit indeference experiment (Chapter 1), attempts to reveal the particle nature of the electrons by determining through which slit the electron has passed, result in the destruction of the interference pattern and, thereby, in the concealment of the wave nature.

Another mode of describing complementarity is to say that the very nature of quantum theory requires us "to regard the space-time co-ordination and the claim of causality, the union of which characterises the classical theories, as complementary but exclusive features of the description, symbolizing the idealisation of observation and definition respectively"[16]. This may be elaborated as follows:

The definition of the state of a physical system presupposes that the system is *closed*; that is, it is free from any external disturbances. The evolution of such a system in time constitutes its causal behaviour and is governed by the dynamical equation of motion. A space-time description of the system, on the other hand, presupposes observation. But observation requires interaction with an external agency of measurement, which, according to the quantum postulate, involves disturbance of the system being observed. In other words, a system, when observed, becomes an *open* system for which, strictly speaking, no 'state' can be defined and for which the laws of causality are no longer applicable. Thus the claim of causality excludes space-time description and vice versa. The simultaneous ('united') use of causal and space-time description in classical mechanics is made possible by the extremely small value of the quantum of action ($\sim 10^{-27}$ erg.sec.) as compared with the actions involved in ordinary sense perceptions.

In the case of radiation, the wave picture, which gives an adequate description of the propagation of light, corresponds to space-time description whereas the particle picture (light quantum), with the associated notions of energy and momentum and their conservation principles, constitutes the causal description.

---

16. Max Jammer (Footnote 8) p. 351

It is obvious from the foregoing that, according to Bohr, complementarity refers to two *modes or pictures of description* which are mutually exclusive but are indispensable for an exhaustive account of quantum phenomena: space-time description is complementary to causal description, wave picture is complementary to particle picture. Pauli[17] gave a slightly different definition of complementarity. He calls two *classical concepts* complementary "if the application of one stands in relation of exclusion to that of the other" in the sense that any experimental set up for measuring the one interferes destructively with any experimental set up for measuring the other. In this sense, position co-ordinate is complementary to the conjugate momentum, time is complementary to energy, and so on.

Bohr's Interpretation of Quantum Mechanics.

As we stated earlier, complementarity and the Uncertainty Principle are the two main planks of Bohr's interpretation of quantum mechanics. The notion of complementarity forms the backbone of the interpretation, while the uncertainty principle ensures that the notion does not lead to any contradictory physical situations. In addition, the following premises also could be regarded as basic to the interpretation:

(1) *However far the phenomena transcend the scope of classical physical explanation, the account of all evidence must be expressed in classical terms.*

This is so because, "by the word 'experiment' we refer to a situation where we can tell others what we have done and what we have learned", so that "the account of the experimental arrangement and of the results of the observations must be expressed in unambiguous language with suitable application of the terminology of classical physics"[18].

(2) *It is impossible to separate sharply the behaviour of atomic objects (i.e. quantum mechanical systems) from their interaction with the measuring instruments which serve to define the conditions under which the phenomena appear.*

Bohr attributes this inseparability of the object of observations from the agency of observation to the individuality or indivisibility, of quantum phenomena arising from the quantum postulate. He contends that this individuality "finds its proper expression in the circumstance that any attempt at subdividing the phenomena will demand a change in the experimental arrangement introducing new possibilities of interaction between objects and measuring instruments which, in principle, cannot be controlled. Consequently, evidence obtained under different experimental arrangements cannot be comprehended within a single

17. Pauli, W. *Handbuch der Physik*, **24** (Springer, Berlin 1933), 2nd ed. 5 p. 126.
18. Bohr, N. (Footnote 8), p. 158

picture, but must be regarded as *complementary* in the sense that only the totality of the phenomena exhausts the possible information about the objects"[19]. It is further stressed that the complementary phenomena appear under mutually exclusive experimental arrangements, so that for the specification of the conditions for any well-defined application of the quantum formalism, the whole experimental arrangements should be taken into account.

(3) *Quantum theory provides an exhaustive description of individual physical phenomena.*

This is to be taken in the sense that the quantum mechanical formalism is an adequate tool for a complementary way of description of physical phenomena. In fact, the quantum mechanical formalism is nothing but a "symbolic scheme permitting only predictions as to results obtainable under conditions specified by means of classical concepts"[20]. The totality of results obtained under mutually exclusive experimental arrangements (that is, complementary situations) exhausts the possible information we can have on a physical system. Moreover, there is no single experimental situation which gives rise to results that are not in conformity with the predictions of the quantum formalism.

For illustration let us consider the double slit interference experiment on electrons described in Section 1.1. As detailed there, the main features of this experiment are the following:

(DS1) If no effort is made to observe the paths of the individual electrons, then the distribution of the electrons on the screen X shows an interference pattern appropriate to waves of wavelength equal to the de Bröglie wavelength of the electrons.

(DS2) If the paths of the electrons are observed, the interference pattern is destroyed.

DS1 and DS2 correspond to mutually exclusive experimental arrangements and, therefore, to complementary features of the properties of electrons. DS1 reveals the wave nature of electrons while DS2 shows their particle-like behaviour. Alternatively, DS1 corresponds to space-time description and DS2 to causal description. "The circumstance that we are presented with a choice of *either* tracing the path of a particle *or* observing interference effects allows us to escape from the paradoxical necessity of concluding that the behaviour of an electron should depend on the presence of a slit through which it could be proved not to pass. We have here to do with a typical example of how the complementary phenomena appear under mutually exclusive experimental arrangements, and are just faced with the impossibility, in the analysis of quantum effects, of drawing any sharp separation between an independent behaviour of atomic objects and

19. Bohr, N. (Footnote 8), p. 159
20. Bohr, N. (Footnote 8), p.159.

their interaction with the measuring instruments which serve to define the conditions under which the phenomena occur"[21]. And it is the uncertainty principle which ensures that we cannot trace the paths of electrons without at the same time leading to the destruction of the interference pattern, as the following analysis shows: Figure 12.1 shows an experimental arrangement for the double-slit diffraction phenomenon, with a device to ascertain the slit through which each

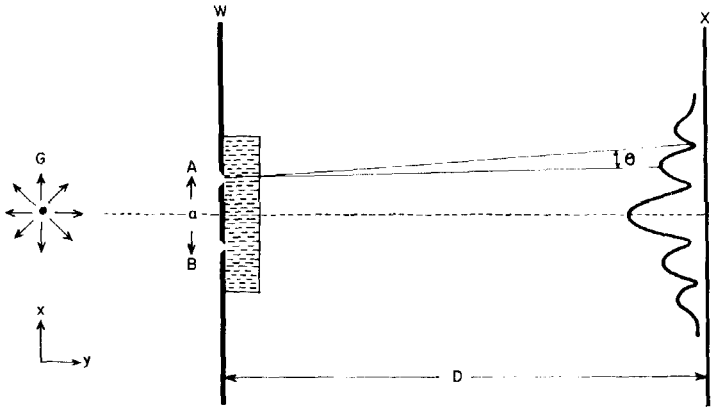


Fig. 12.1 Double-Slit experiment with a device to observe the electrons just after they pass the slits.

electron passes before it reaches the screen X. The detection procedure would involve a transfer of momentum ΔP_x to the electrons in the plane of the slits. If the diffraction pattern is not to be destroyed in the process, ΔP_x should be substantially smaller than the momentum required to throw an electron from a maximum of the diffraction pattern to a neighbouring minimum. If P_y is the momentum of the electron, then, referring to Fig. 12.1 we should have,

$$\Delta P_x \ll P_y \theta. \tag{12.2}$$

But, from wave theory (wave optics) we have,

$$\theta = \lambda/(2a) \tag{12.3}$$

where λ is the de-Bröglie wave length of the electron and 'a' is the distance between the slits, and where $D >> a >> \lambda$, D being the distance between W and X. At the same time, in order to be able to tell through which slit the electron has passed, its position in the x-direction should be determined to an accuracy Δx which is better than half the distance between the slits. That is,

$$\Delta x < a/2. \tag{12.4}$$

From Eqs. (12.2 - 4) we see that the condition for us to be able to determine through which slit each electron passes, without destroying the interference pattern, is that,

21. Bohr, N. (Footnote 8), p. 163 - 164.

$$\Delta x.\Delta P_x \ll h/4. \tag{12.5}$$

But this condition violates the uncertainty relationship (3.29a) and is, therefore, not realisable, implying that any successful attempt at determining the slit through which each electron passes before arriving at the screen, X, would lead to the destruction of the interference pattern.

Fig. 1.1 could be regarded as an experimental arrangement for measuring the x-component of the momentum of the electrons (since the distribution of intensity in the interference pattern determines, to a good approximation, the angular distribution of the electrons leaving the slits), while Fig. 12.1 is an arrangement for measuring their x-position. Thus, according to Pauli's definition of complementarity, Fig. 1.1 (the arrangement) interferes destructively with Fig. 12.1 and *vice versa*.

We will postpone to Section 12.4 a discussion of Bohr's explanation of the EPR paradox.

12.3 THE ENSEMBLE INTERPRETATION

The Ensemble Interpretation was proposed by Albert Einstein[22] as an alternative to the Copenhagen Interpretation which he considered quite unsatisfactory. In fact, Einstein looked upon the Copenhagen Interpretation as nothing but a 'tranquilizing philosophy' as the following passage from a letter he wrote to Schrödinger in 1928 reveals:

> "The Heisenberg-Bohr tranquilising philosophy is so delicately contrived that, for the time being, it provides a gentle pillow for the true believer from which he cannot very easily be aroused"[23].

On a more specific level, Einstein's objection to the Copenhagen interpretation is on its insistence that quantum theory provides a *complete* description of an *individual physical system.* Thus, in his "Reply to Criticisms" in the volume of essays[24] presented on his seventieth birthday, he writes (p. 671):

> "One arrives at very implausible theoretical conceptions, if one attempts to maintain the thesis that the statistical quantum theory is in principle capable of producing a complete description of an individual physical system. On the other hand, those difficulties of theoretical interpretation disappear, if one views the quantum mechanical description as the description of ensemble of systems".

---

22. The account of Einstein's interpretation of quantum mechanics given here is based on the following references: (i) Einstein, A. and Infled, *The Evolution of Physics* (Simon and Schuster, New York 1938, 4th Printing, 1961), pp. 280-294; (ii) Ballentine, L.E. *Rev. Mod. Phys.* **42**, 358 (1970) and (iii) Ballentine, L.E; *Amer. J. Phys.* **40**, 1763 (1972).
23. Letter to Schrödinger, dated May 31, 1928 in *Letters on Wave Mechanics,* (ed) K.Przibram, translated by M.J. Klein (Philosophical Library, New York 1967), p. 31.
24. *Albert Einstein : Philosopher - Scientist,* (ed) P.A.Schilpp (Library of the Living Philosophers, Evanston, Illinois 1949; reprinted by Harper and Row, New York 1959).

In order to illustrate and establish this contention, Einstein proposed and cited a number of thought-experiments over a period of years starting from the Fifth Solvay Conference in 1927. At this conference held in Brussels from 24 to 29 October he considered the situation pictured in Fig. 12.2. S is a diaphragm with a small opening O at the centre; P represents an electron-sensitive screen in the form of a hemisphere. A beam of electrons falls on S. Some of the electrons will pass through O. Because of the smallness of O the electrons will be diffracted at O, according to quantum mechanics. The wave function of the diffracted electrons will be represented by a spherical wave which will have non-vanishing value on the whole of P. In particular, the wave-function will have nonzero values both at A and B on P.

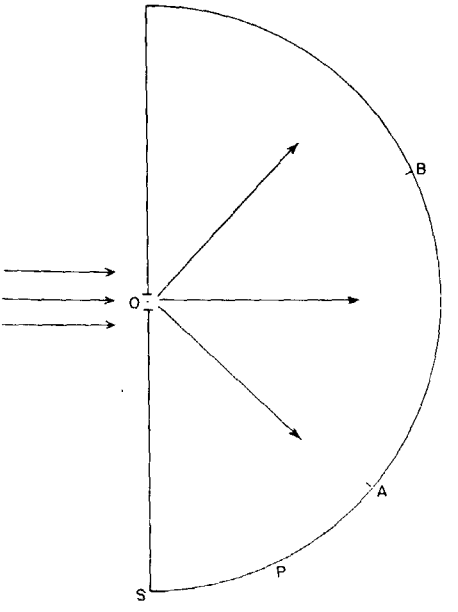


Fig. 12.2 Einstein's thought-experiment on electron diffraction by a single slit.

Now, if an electron is observed at A, we immediately know that it is not at B. Einstein then distinguished two different points of view:

I1. The wave function does not correspond to a single electron, but to a cloud of electrons extended in space. The theory does not give any information about the individual processes, but only about the ensemble of an infinity of elementary processes.

I2. The theory has the pretention to be a complete theory of individual processes.

According to I1, $|\psi|^2$ expresses (in a statistical sense) the probability for an electron of the cloud to exist at a particular place. In other words, $|\psi|^2$ will give the *distribution* of the electrons of the cloud over P. Conversely, in order to determine ψ, one has to observe the distribution of the electrons over the whole of P and not just at one location A.

According to I2, on the other hand, $|\psi|^2$ expresses the probability that at a certain instant *one and the same electron* be found at the different places on P. As a result, observing the electron at A results in the vanishing of the ψ everywhere else on P (*reduction of the wave function*). Since ψ has finite value at B until the electron is observed at A, the instantaneous vanishing of ψ at B requires a peculiar action-at-a distance between A and B, contrary to the principles of relativity. So Einstein concluded: "In my opinion, one can only remove this objection in this manner, that one does not describe only the process by the Schrödinger wave, but *at the same time one localizes the particle during the propagation*". *Since the* description by means of the ψ-function does not contain such a localization, the description of an individual system in terms of the ψ-function is incomplete.

Schrödinger's Cat

Another example which illustrates the difficulties inherent in the Copenhagen viewpoint that the wave function provides a complete description of an individual system, is the following argument advanced by Schrödinger in a review article[25]: Imagine a chamber which houses a cat together with a machine, a bottle of cyanide and a radioactive substance. The machine has a triggering mechanism which, when activated by the decay of an atom of the radioactive substance, will trip a hammer which will break the bottle of cyanide. The quantity of the radioactive substance in the chamber is such that there is equal probability for one atom to decay in an hour and for none to decay during that time. If one describes this entire system according to quantum theory, then at the end of one hour the wave function of the system would be a linear combination of equal parts of functions corresponding to a live cat and a dead cat.

If one adopts the viewpoint that the wave function furnishes a complete description of an individual system, one has to conclude that, at the end of the hour, the cat is neither dead nor alive, just as in the example of the electron diffraction the electron has no definite position but is (potentially) present all over P. But if one looks to see whether the cat is actually dead or alive and finds that it is really dead, one has to assume that it is the act of looking that killed the cat (reduction of the wave function)!

According to the Ensemble Interpretation, on the other hand, the description in terms of the wave function refers to an ensemble of a large number of "Schrödinger's cats". That the wave function is a linear combination of equal

25. Schrödinger, E. *Naturewiss*, 28, 807 (1935).

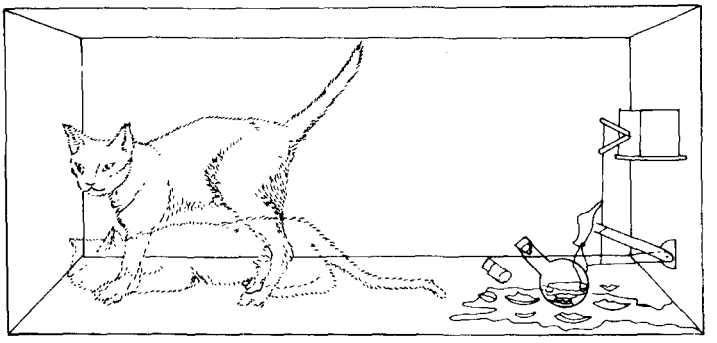


Fig. 12.3 The Schrödinger's Cat experiment.

parts of functions corresponding to live and dead cats means that, in about half of the cases the cats will be dead (signifying that the atom has decayed) and in the other half cases the cats will be alive (the atom has not decayed).

Superposition of States

The Ensemble Interpretation of quantum mechanics was expounded by Einstein for the first time in an article entitled "Physics and Reality"[26]. In this article he presented the following argument in support of his view that the ψ function does not describe an individual system:

A system is initially in its state of lowest energy ϵ_1. The corresponding wave function is ψ_1. The system is now subjected to a small time-dependent perturbation for a finite interval of time. As a result, the wave function of the system takes the form

$$\psi = \sum_r c_r \psi_r, \tag{12.6}$$

where the different ψ_r's represent the stationary states of the system and the coefficients c_r satisfy the condition,

$$\sum_r |c_r|^2 = 1 \tag{12.7}$$

"Does ψ describe a real state of the system? If the answer is yes, then we can hardly do otherwise than to ascribe to this state a definite energy ϵ, and, in particular, an energy which exceeds ϵ_1 by a small amount (in any case $\epsilon_1 < \epsilon < \epsilon_2$)." But the experiments of Franck and Hertz[27] on electron impact show that an individual system can have only one of the discrete energies $\epsilon_1, \epsilon_2, \ldots, \epsilon_r, \ldots$. Einstein is, thus led to the conclusion: "It seems to be clear,

26. Einstein, A. *J. Franklin Institute*, **221**, 349 (1936).
27. Franck, J. and Hertz, G. *Physikalische Zeitzchrift*, **17**, 409 (1916); **20**, 132 (1919).

therefore, Born's statistical interpretation of quantum theory is the only possible one. The ψ-function does not in any way describe a state which could be that of a single system; it refers rather to many systems, to an ensemble of systems' in the sense of statistical mechanics. If, except for certain special cases, the ψ-function furnishes only *statistical* data concerning measurable magnitudes, the reason lies not only in the fact that the *operation of measuring* introduces unknown elements, which can be grasped only statistically, but because of the fact, the ψ-function does not, in any sense, describe the state of one single system''.

The absolute square of c_r in Eq. (12.6) gives the probability for the system to be in the state ψ_r with energy ϵ_r. In other words, $|c_r|^2$ is proportional to the number of individual systems in the ensemble which have energy ϵ_r.

This ensemble interpretation is also described in detail in the book by Einstein and Infeld [Footnote 22(i)]. We give below a few quotations from this book, which give a clear picture of the nature of Einstein's Interpretation of quantum mechanics:

> ''The laws of quantum physics are of a statistical character. This means: they concern not one single system but an aggregation of identical systems; they cannot be verified by measurement of one individual, but only by a series of repeated measurements'' (p. 284).

After referring to phenomena like radioactivity where the laws of quantum physics work well, the authors say:

> ''The theory works splendidly because all these phenomena involve large aggregations and not single individuals'' (p. 285).

They contrast the statistical nature of quantum theory with the statistical nature of certain classical theories like that of population statistics and the kinetic theory of gases in which no predictions can be made for individual cases but the average behaviour in a large number of cases can be predicted fairly accurately. They point out that the similarity between quantum theory and those classical theories lies chiefly in their statistical character, but the differences are equally important. In the case of population,

> ''Our statistical view is gained by the knowledge of individual cases. Similarly, in the kinetic theory of matter, we have statistical laws governing the aggregation, gained on the basis of the individual laws.
> ''But in quantum physics the state of affairs is entirely different. Here the statistical laws are given immediately. The individual laws are discarded'' (p. 286).

We have already considered, in Section 1.1, the explanation of the double-slit interference experiment on the basis of the Ensemble Interpretation. We will discuss the explanation of the EPR paradox in the next section.

12.4 EXPLANATIONS OF THE EPR PARADOX

Both Bohr and Einstein have offered explanations of the EPR paradox, based on their respective interpretations of the quantum formalism.

Bohr's Explanation

Bohr's explanation[28] is essentially that $S_x^{(1)}$ and $S_z^{(1)}$ (and, hence, also $S_x^{(2)}$ and $S_z^{(2)}$ deduced from $S_x^{(1)}$ and $S_z^{(1)}$ respectively) are not simultaneously elements of physical reality but are complementary attributes of the particle. For, measurements of $S_x^{(1)}$ and $S_z^{(1)}$ require different experimental set ups (different orientations of the Stern-Gerlach magnet) and, according to the principle of Complementarity (Section 12.2), the result of an observation cannot be divorced from the experimental set up which yielded the result; it is the result and the experimental set up jointly that define a quantum process. Bohr also contends that the definition of physical reality as given in (EPR.2) in Section 12.1 is defective in the phrase 'without in any way *disturbing* a system'. Bohr suggests its replacement with the phrase 'without in any way *influencing* a system' since measurement on particle number 1, though does not disturb particle number 2, has nevertheless "an influence on the very conditions that define the possible types of predictions regarding the future behaviour"[29] of particle number 2. As a result of these considerations, no more than one component of $S^{(2)}$ can be regarded as an element of physical reality in any particular experiment.[30]

Einstein's Explanation

According to Einstein's interpretation of quantum mechanics, ψ_0 given by Eq. (12.1) represents an ensemble of two-particle systems in the singlet state rather than a single such system. Therefore the EPR experiment, referring as it does to an individual pair, is irrelevant as far as ψ_0 (as well as quantum mechanics) is concerned. ψ_0 can be used to answer only questions of a statistical nature concerning the ensemble. However, such answers require a series of repeated measurements rather than a single measurement on an individual system. For example, the absolute square of the coefficient of $\chi_{1/2}^{(1)}(\mathbf{a})\,\chi_{-1/2}^{(2)}(\mathbf{a})$ in Eq. (12.1) denotes the number of pairs in the ensemble with spin-component parallel to **a** for particle number 1 and antiparallel to **a** for particle number 2. Eq.(12.1) shows that this proportion is the same as the proportion of the number of pairs in the ensemble with spin-component antiparallel to **a** for particle number 1 and parallel

28. Bohr, N. *Phys. Rev.* **48**, 696 (1935)
29. Bohr, N. (Footnote 8)
30. In the language of the Hilbert space, Bohr's explanation refers to the fact that the angular momentum can have different *representations*, but that in any one representation *only one* of the components can be diagonal.

to **a** for particle number 2. But in order to verify this fact, one will have to make repeated observations on the pairs constituting the ensemble. Thus one could set up two Stern-Gerlach magnets on opposite sides of the ensemble so as to measure the spin-components (along a specified direction **b**) of the two particles of a pair which has decayed. We will label the particle which comes to the left-side detector (DL) as 1 and the one which goes to the right-side detector (DR) as 2. When the observation is repeated for a large number of pairs we would get a *certain number* for which DL has recorded spin-up while DR has recorded spin-down and about an *equal number* for which DL has recorded spin-down and DR spin-up. "This correlation between the spins of the particles will be the same no matter which component is measured"[31].

It is, however, essential for the validity of this ensemble interpretation that every recording of spin-up (spin-down) at DL should *not* be accompanied by a recording of spin-down (spin-up) at DR. For, if it were so, the spin-correlation specified by ψ_0 would have been applicable to each and every pair on which observation is made and, therefore, also to an isolated pair. But Einstein is emphatic that "quantum physics deals only with aggregations and its laws are for crowds and not for individuals"[32]. Therefore, there should be cases where both DL and DR record spin-up states and both record spin-down states. We cannot predict the proportion of these states on the basis of ψ_0 alone. In order to obtain these proportions, one will have to find out the eigenvectors of the operator $(\hat{\sigma}_1 \cdot \mathbf{b})\,(\hat{\sigma}_2 \cdot \mathbf{b})$ in the representation spanned by the basis vectors $X^{(1)}_{m_1}(\mathbf{a})\,X^{(2)}_{m_2}(\mathbf{a})$, where $m_1 = \pm\frac{1}{2}$, $m_2 = \pm\frac{1}{2}$ and where $\mathbf{a} \neq \mathbf{b}$. There would be four linearly independent eigenvectors, two each belonging to the degenerate eigenvalues +1 (corresponding to spin up-spin up and spin down-spin down results) and − 1 (representing spin down-spin up and spin up-spin down results). The wave functions are found[33] to be consistent with the conclusion arrived at by Mermin[7] in a *gedanken* experiment supposed to represent the experimental findings of Alain Aspect and his group at the University of Paris, namely that each of the four cases occurs an equal number of times and constitutes one-fourth of the total.

An Alternative Explanation

It is possible to have a third explanation[34] based purely on the mathematical structure (the Hilbert space formulation) of quantum mechanics described in chapter 3. According to the Postulates I and II stated there, it follows that the results of the measurement of a physical observable is represented by the eigenvalue equation of the corresponding Hermitian operator. It also implies that if we try to measure the value of a physical observable in a state which is not an

---

31. Ballentine, L.E. (Footnote 22(ii), p. 364)
32. Einstein, A. and Infeld, 4. (Foot note 22(i), p. 286).
33. Menon, R.K. and Thankappan, V.K. (to be published)
34. Thankappan, V.K. and Menon, R.K., Proceedings of the 5th Asia Pacific Physics Conference, Kuala Lumpur, August 1992 (in press).

eigenstate (eigenvector) of the Hermitian operator to start with, then the measurement process (that is, the interaction between the measuring instrument and the physical system) will *carry the system over to an eigenstate of the operator.* Then the measured value of the observable will refer to the state of the system at the end of the measurement process, which need not be the state the system was in at the beginning, or before, the measurement.

The above feature of quantum mechanics necessitates the following modifications in the EPR argument:

If $\psi_0(1, 2)$ of Eq. (12.1) represents the state of the pair of particles before the measurement was made (that is, at the time of decay), then the measurement process will change the state of the pair to $\phi(1, 2)$, where,

$$\phi(1,2) = \chi_m^{(1)}(\mathbf{b})\,\chi_{-m}^{(2)}(\mathbf{a}). \tag{12.8}$$

Here, $m = \pm\frac{1}{2}$ and **b** is the unit vector along which the spin of particle number 1 is measured; that is, **b** specifies the axis of the Stern-Gerlach apparatus. The reasoning behind this change is as follows:

$\psi_0(1, 2)$ is an eigenvector of $\hat{S}_a = \hat{S}_a^{(1)} + \hat{S}_a^{(2)}$, the operator corresponding to the spin of the *pair* along **a**, but is not an eigenvector of either $\hat{S}_b^{(1)}$ (or even $\hat{S}_a^{(1)}$) or $\hat{S}_a^{(2)}$, the operators representing the components of the spins of the individual particles. However, quantum mechanics stipulates that, at the end of a measurement process designed to measure the component of the spin of particle number 1 along **b**, the state of the system must be an eigenvector of $\hat{S}_b^{(1)}$.

In the theory of angular momentum coupling, ψ_0 is a vector in the *coupled* representation whereas ϕ is a vector in the *uncoupled* representation (see Section 5.5A). In the uncoupled state, the spin components of the two particles are not correlated; for a given value of $S_b^{(1)}$, $S_a^{(2)}$ could be either $+(\hbar/2)$ or $-(\hbar/2)$, that is, the sum of $S_b^{(1)}$ and $S_a^{(2)}$ need not be zero or any other *fixed* value. Therefore, the EPR assumption that the value of $S_a^{(2)}$ can be deduced from the measured value of $S_b^{(1)}$ has no basis in quantum mechanics. Of course, this is not due to any violation of the conservation law relating to angular momentum, but is due to the fact that the role of the measuring instrument cannot be ignored in balancing the angular momentum. A closer look at the measurement process will elucidate this point further:

At the time of decay of the pair, let the spins of the two particles be quantized along the unit vector **a** such that $S_a^{(1)} = +\hbar/2$ and $S_a^{(2)} = -\hbar/2$ (Fig. 12.4). This means that the spin vector of particle number 1 would be precessing around **a** such that the projection of $\mathbf{S}^{(1)}$ along **a** is equal to $(\hbar/2)$. The spin of particle number 1 is now measured with a Stern-Gerlach apparatus whose axis is along **b**. This measurement process will force a reorientation of the spin vector of particle number 1 such that it now precesses with a fixed value of its projection along **b** instead of along **a**. This change comes about as a result of the (local) interaction (involving

exchange of angular momentum) between particle number 1 and the Stern-Gerlach apparatus and not as a result of any 'spooky actions at a distance' between the two particles. We may now distinguish between two cases:

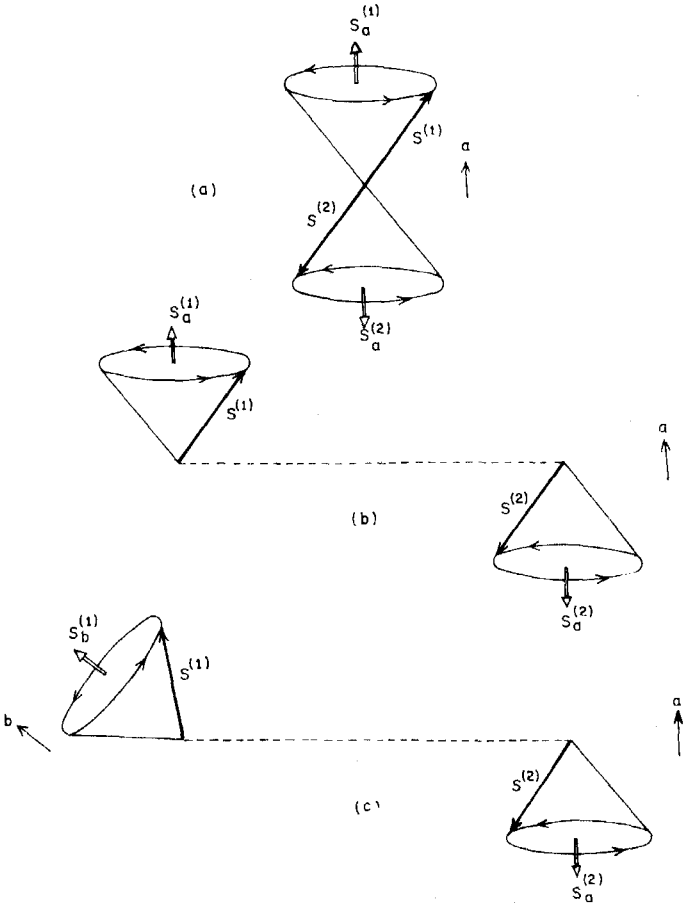


Fig. 12.4 The EPR experiment. The diagram shows the spins of the two particles at the time of decay but before separation (a); after separation but before measurement (b); and after measurement(c).

Case 1 : $|\ (\mathbf{a}\cdot\mathbf{b})\ | < 1$

For this case quantum mechanics predicts that $S_b^{(1)} = \pm(\hbar/2)$. But quantum mechanics also predicts the same two possible values of $S_b^{(1)}$ even when $S_a^{(1)} = -(\hbar/2)$ instead of $+(\hbar/2)$. Therefore, it is not possible to deduce the value of

$S_a^{(1)}$ (and from it the value of $S_a^{(2)}$ using the relation $S_a^{(1)} + S_a^{(2)} = 0$) from the measured value of $S_b^{(1)}$. The basic cause of this disability is the quantization of angular momentum which (i) restricts the observable values of the spin along any direction whatsoever to just the same two values $\pm(\hbar/2)$ and (ii) makes it impossible to predict beforehand which of the two values will result in any particular measurement as the change in the orientation of the spin (say from precessing around **a** to precessing around **b**) does not take place continuously but takes place in discrete jumps. In the corresponding classical case, the change in the angular momentum resulting from the interaction between the particle and the Stern-Gerlach magnet would be both continuous and predictable.

Case 2 : $|\ (\mathbf{a} \cdot \mathbf{b})\ | = 1$

This is the case when the choice of the axis of the Stern-Gerlach apparatus happens to be the same as the quantization axis of the pair at the time of decay. According to quantum mechanics, the measurement process will cause no change in the orientation of the spin of particle number 1, so that $S_b^{(1)} = S_a^{(1)} = -S_a^{(2)}$. Thus $S_a^{(2)}$ can be deduced. In reality, however, even in this case, it will not be possible to deduce the value of $S_a^{(2)}$ because there is no way of ascertaining that, in fact **b** is the same as **a**[35]. But even if we assume, for arguments sake that **b** is the same as **a**, we can deduce only one component of $\mathbf{S}^{(2)}$, namely $S_a^{(2)}$ by measurement done on particle number 1. Any other choice of **b** would mean that $|(\mathbf{a} \cdot \mathbf{b})| < 1$ which is the same as case 1.

In short, we can at most deduce only one component of $\mathbf{S}^{(2)}$ from measurement done on $\mathbf{S}^{(1)}$. Therefore, there is no circumstance which leads to a violation of Eq. (10.35a) and there is no paradox.

12.5 THE HIDDEN VARIABLE THEORIES

A bye-product of the interpretational dispute on quantum mechanics between Bohr and Einstein was the development of a class of theories known as the Hidden Variable Theories (HVT). The viewpoint adopted by Einstein with regard to the interpretational problem and the Ensemble Interpretation advocated by him (Section 12.3), consider quantum mechanics as defective or unsatisfactory in certain respects. The unsatisfactory feature was illuminated clearly in the EPR paradox (Section 12.1) and is contained in the Einstein's theorem formulated in connection with it. As explained there, we have to regard quantum mechanics either as an *incomplete* theory or as a *non-local* theory. One could try to remedy

35. Remember that in the EPR experiment, there is only one pair. If there are many pairs as in an ensemble, we could assume that for some of the pairs the axis of quantization would be the same as **b**.

the defect by inventing a theory in which the *microstate* of a particle is characterised by one or more parameters ('hidden variables') λ in addition to the ψ-function. In such a theory, ψ would describe a *macrostate* obtained by averaging over the microstates (that is, over λ).

The development of hidden variable theories has been inhibited for nearly two decades by a theorem due to von Neumann[36] (which purportedly proved the impossibility of such theories unless the predictions of quantum theory are factually wrong), until Bohm showed[37] that a true HVT can be developed in spite of the theorem and Bell[38] established the irrelevance of the theorem for realistic hidden-variable theories.

The hidden variable theories could be broadly classified as theories of the *first kind* and as theories of the *second kind*[39]. Theories of the first kind consider incompleteness as the basic defect of quantum theory. Such theories would yield quantum theory as a statistical equilibrium limit of (or, as an average over) the microstates, and are therefore *compatible* with quantum theory. Experimental verification of the predicted deviations of these theories from quantum theory is, however, made difficult or impossible by the extremely short time ($=10^{-14}$ sec.) within which the microstates reach statistical equilibrium[37].

Theories of the second kind, on the other hand, are designed to eliminate the *non-local* feature of quantum mechanics (these are, therefore, called *local* HVT). These theories are rivals to quantum mechanics rather than forming its microscopic basis. As shown by Bell[38], the question whether such a local hidden variable theory or quantum theory represents the laws of Nature is one that could be settled by experiments—the former will satisfy *Bell's inequalities* [(12.28) below] whereas the latter will violate them.

It may be emphasized here that, though the hidden variable theories are logical corollaries of Einstein's attitude towards quantum theory as reflected in the analysis of the EPR experiment, Einstein himself was not enthusiastic about these theories. In fact, when Bohm's paper[39] on HVT appeared, Einstein wrote to Born[40]: "Have you noticed that Bohm believes (as de Bröglie did, by the way, 25 years ago) that he is able to interpret quantum theory in deterministic terms? That way seems too cheap to me". It is obvious that Einstein accepted the statistical quantum theory as correct in its domain of applicability and did not consider lack of determinism as its major defect. In his view, incompleteness (the fact that the

---

36. von Neumann, J. *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin 1932); translation: *The Mathematical Foundation of Quantum Mechanics* (Princeton University Press, 1955).
37. Bohm, D. *Phys. Rev.* **85**, 166 and 180 (1952).
38. Bell, J.S. *Physics*, **1**, 195 (1964).
39. This classification of hidden variable theories is due to Belinfante, F.J. [see his book: *A Survey of Hidden-Variables Theories* (Pergamon Press, Oxford 1973)]; Belinfante also describes another class of HVT, namely, the *zeroth kind*: these are the ones which are prohibited by von Neumann's theorem.
40. *The Born-Einstein Letters* (Walter and Company, New York, 1971), Letter No. 99 dated 12 May 1952.

ψ-function does not describe an individual system) was the defect to be removed by a theory more general than quantum theory. However, he did not believe that this more general theory has quantum theory as its starting point.

In this book, we do not propose to go into the details of specific hidden variable theories[41]. Instead we will give a brief account of some of the theorems, arguments and conclusions which have been crucial for the development, understanding and establishment of hidden-variable theories.

von Neumann's Theorm

This theorem states[42]: ".....the present system of quantum mechanics would have to be objectively false, in order that another description of the elementary processes than the statistical one be possible".

This is a conclusion which has effectively deterred the development of hidden-variable theories for nearly two decades. It would, therefore, be of interest to present the proof if only to point out the loopholes in it later.

The proof of the theorem consists of two parts. In part one, it is established that there exists an unique Hermitian statistical operator $\hat{\rho}$ such that for any arbitrary Hermitian operator $\hat{A}$,

$$\langle \hat{A} \rangle = \mathrm{Tr}\ (\hat{\rho}\hat{A}) \tag{12.9}$$

where <...> represents the expectation value. In the second part, it is shown that, in the case of a dispersionless ensemble, $\hat{\rho}$ is either a null or a unit operator.

The proof of part one is based on the following assumptions:

(A1) The correspondence between Hermitian operators and physical observables is one to one. In other words, every Hermitian operator corresponds to a physical observable.

(A2) If two observables are represented by operators $\hat{A}_1$ and $\hat{A}_2$, then the sum of the two observables is represented by the operator $\hat{A}_1 + \hat{A}_2$.

(A3) If $\hat{A}_1$ and $\hat{A}_2$ are arbitrary observables and a_1 and a_2 are arbitrary real numbers, then,

$$\langle a_1\hat{A}_1 + a_2\hat{A}_2 \rangle = a_1\langle \hat{A}_1 \rangle + a_2\langle \hat{A}_2 \rangle, \tag{12.10}$$

for all possible ensembles (states), irrespective of whether $\hat{A}_1$ and $\hat{A}_2$ are commuting or not.

(A4) If the observable $\hat{A}$ is nonnegative (that is, positive semidefinite), then, $\langle \hat{A} \rangle \geq 0$.

Let $\hat{A}$ be an arbitrary Hermitian operator. It can be written as

41. Such details can be found in the book by Belinfante (Footnote 39).
42. Ref. 36 (English translation) p. 325.

$$\hat{A} = \sum_{m,n} |n\rangle A_{nm} \langle m|$$

$$= \sum_{n} \hat{U}^{(n)} A_{nn} + \sum_{\substack{m,n \\ (m<n)}} \{\hat{V}^{(nm)} \mathrm{Re}[A_{mn}] + \hat{W}^{(nm)} \mathrm{Im}[A_{mn}]\}, \tag{12.11}$$

where, $$A_{nm} = A_{mn}^{*} = \langle n | A | m \rangle, \tag{12.12}$$

and $$\hat{U}^{(n)} = |n\rangle\langle n|, \tag{12.13a}$$

$$\hat{V}^{(nm)} = |n\rangle\langle m| + |m\rangle\langle n| \tag{12.13b}$$

$$\hat{W}^{(nm)} = -i(|n\rangle\langle m| - |m\rangle\langle n|). \tag{12.13c}$$

In (12.11). Re [ ] and Im [ ] represent respectively the real and the imaginary parts.

(12.13 a-c) are Hermitian operators and, according to assumption (A1), are observables. Applying (A2) and (A3), then we have,

$$\langle \hat{A} \rangle = \sum_{n} A_{nn} \langle \hat{U}^{(n)} \rangle$$

$$= \sum_{\substack{mn \\ (m<n)}} \{\mathrm{Re}[A_{mn}] \langle \hat{V}^{(nm)} \rangle + \mathrm{Im}[A_{mn}] \langle \hat{W}^{(nm)} \rangle\}, \tag{12.14}$$

or,

$$\langle \hat{A} \rangle = \sum_{m,n} \rho_{nm} A_{mn} = \mathrm{Tr}(\hat{\rho} A), \tag{12.15}$$

where $\hat{\rho}$ is defined by

$$\rho_{nn} = \langle U^{(n)} \rangle, \tag{12.16a}$$

$$\rho_{mn} = \frac{1}{2}\{\langle V^{(nm)} \rangle + i\langle W^{(nm)} \rangle\}, \qquad (m < n), \tag{12.16b}$$

$$\rho_{nm} = \frac{1}{2}\{\langle V^{(nm)} \rangle - i\langle W^{(nm)} \rangle\}, \qquad (m < n). \tag{12.16c}$$

From (12.16b) and (12.16c) we see (since expectation values of observables are real) that

$$\rho_{mn} = (\rho_{nm})^{*}, \qquad \text{or} \qquad \hat{\rho}^{\dagger} = \hat{\rho}. \tag{12.17}$$

Thus, $\hat{\rho}$ is Hermitian. This completes the first part of the proof of von Neumann's theorem. It can also be shown that $\langle \phi | \hat{\rho} | \phi \rangle \geq 0$, for all ϕ, so that $\hat{\rho}$ is non-negative definite, according to (A4).

The second part of the proof makes use of an additional assumption:

(A5) If an observable is represented by the operator $\hat{A}$, then a function $\hat{f}$ of that observable is represented by $\hat{f}(\hat{A})$.

In particular, therefore, the square of the observable represented by $\hat{A}$, is represented by $\hat{A}^2$. Then, in the case of a dispersionless ensemble,

$$\langle(\hat{A} - \langle\hat{A}\rangle)^2\rangle = 0, \tag{12.18a}$$

or,
$$\langle\hat{A}^2\rangle = \langle\hat{A}\rangle^2, \tag{12.18b}$$

for all observables $\hat{A}$, since the expectation value is equal to an eigenvalue in the case of such an ensemble.

Substituting from (12.15) in (12.18b), we have,

$$\mathrm{Tr}\,(\hat{\rho}\hat{A}^2) = [\mathrm{Tr}\,(\hat{\rho}\hat{A})]^2. \tag{12.19}$$

Choosing $\hat{A}$ to be the projection operator $\hat{P}$, where,

$$\hat{P} = |\phi\rangle\langle\phi| \text{ with } \langle\phi|\phi\rangle = 1, \text{ we get,}$$

since $\hat{P}^2 = \hat{P}$ [see Eq. (2.69a)],

$$\mathrm{Tr}\,(\hat{\rho}P)\,[\mathrm{Tr}\,(\hat{\rho}\hat{P}) - \hat{1}] = \hat{0}. \tag{12.20}$$

That is,

$$\mathrm{Tr}\,(\hat{\rho}\hat{P}) = \hat{0} \text{ or } \hat{1}$$

or,

$$\langle\phi|\hat{\rho}|\phi\rangle = 0 \text{ or } 1, \tag{12.21}$$

since
$$\hat{P}\,|m\rangle = \delta_{m,\phi}\,|\phi\rangle$$

But $|\phi\rangle$ is arbitrary. Therefore, (12.21) can be satisfied only if

$$\hat{\rho} = \hat{0}, \text{ or } \hat{\rho} = \hat{1}. \tag{12.22}$$

That is, for dispersion-free ensembles, the statistical operator is either null or unity.

Now, $\hat{\rho} = \hat{0}$ implies that $\langle\hat{A}\rangle = 0$ for all $\hat{A}$, which is an unacceptable result. The case $\hat{\rho} = \hat{1}$ gives $\langle\hat{A}\rangle = \mathrm{Tr}\,(\hat{A})$, which cannot be true except for the one dimensional vector space (where A is an 1×1 matrix). Thus the assumption of a dispersion-free ensemble leads to unacceptable conclusions. von Neumann, therefore, argues that, provided his assumptions are accepted, there are no dispersion-free ensembles and hence there cannot be any hidden variables. For, a hidden variables state, by definition, is dispersion-free as every observable has an unique value in such a state. Thus the conclusion: "It is therefore not, as is often assumed, a question of reinterpretation of quantum mechanics,—the present system of quantum mechanics ....." [42] appears well-founded.

Bell's Rebuttal of von Neumann's Proof

The credit for pinpointing the defect in von Neuman's proof of his theorem goes to Bell[43], though the irrelevance of the theorem for the hidden variable question in quantum theory was indicated also by Bohm's hidden-variables theory introduced in 1952[37]

---

43. Bell, J.S. *Revs. Mod. Phys.* **38**, 447 (1966).

As we have emphasized, an important part of von Neumann's proof is the result that any ensemble can be characterised by an unique statistical operator, whether the ensemble corresponds to a quantum mechanical state described by the ψ-function or to a hidden-variables state described by ψ and λ. However, the fact is that characterisation by means of a statistical operator is not valid for a hidden-variables state[44]. Therefore, at least one of the assumptions (A1) to (A4) made by von Neumann must be false for the case of a hidden-variables state. Bell pointed out that it is actually assumption (A3) that is at fault. This assumption states that the average of the sum of two observables is equal to the sum of the averages of the two observables separately. Though this is valid for all quantum mechanical states, it is a nontrivial property as far as hidden-variables states are concerned. For, in a hidden-variables state, expectation values are eigenvalues of the corresponding operators, and it is well known that eigenvalues of non-commuting operators do not have the additivity property (12.10) postulated by (A3). The physical basis of this would be clear from a consideration of the case where the observables are the spin components of a spin 1/2 particle. Putting $\hat{A}_1 = \hat{\sigma}_x$, $\hat{A}_2 = \hat{\sigma}_y$ and $a_1 = a_2 = 1$, Eq. (12.10) reduce to

$$\langle \hat{\sigma}_x + \hat{\sigma}_y \rangle = \langle \hat{\sigma}_x \rangle + \langle \hat{\sigma}_y \rangle. \tag{12.23}$$

For a dispersion-free (hidden-variables) state, (12.23) states that the eigenvalue of $(\hat{\sigma}_x + \hat{\sigma}_y)$ is the sum of the eigenvalues of $\hat{\sigma}_x$ and $\hat{\sigma}_y$. But the eigenvalues of $(\hat{\sigma}_x + \hat{\sigma}_y)$ are $\pm\sqrt{2}$ whereas the eigenvalues of $\hat{\sigma}_x$ and $\hat{\sigma}_y$ are ± 1, so that Eq. (12.23) is not satisfied. The reason is that the measurements of $\hat{\sigma}_x$, $\hat{\sigma}_y$ and $(\hat{\sigma}_x + \hat{\sigma}_y)$ require three different orientations of the Stern-Gerlach magnets.

Thus, von Neumann's mistake was in assuming that what is true for quantum theory is also true for a hidden-variables theory. To put it differently, von Neumann's impossibility proof is applicable to only a particular class of hidden-variables theories (what Belinfante calls "the zeroth kind") and not to realistic hidden-variables theories (of the first and of the second kinds).

Gleason's Work

Gleason's[45] was able to establish the existence of an unique, self-adjoint, non-negative statistical operator satisfying Eq. (12.9) using von Neumann's assumption (A3) for only *commuting operators* for separable real or complex Hilbert spaces of dimension greater or equal to 3. Thus, objections raised in connection with von Neumann's proof are not valid here. Therefore, for Hilbert spaces of dimension greater or equal to 3, the second part of von Neumann's proof and the conclusion regarding the impossibility of hidden-variables theories remain valid.

However, in proving his result Gleason makes use of the assumption that the result of a measurement of an observable A is independent of what other

---

44. A proof is given in the book by Belinfante, F. J. (Footnote 39), section 2.2.
45. Gleason, A.M. *J. Math. Mech.* **6**, 885 (1957)

compatible observables are simultaneously measured. As Bell has shown[43], once this assumption is removed and a measurement result is allowed to depend on the *whole experimental arrangement,* it is possible to introduce hidden-variables.

The work of Kochen and Specker[46] on the impossibility of hidden-variables theories is based on an assumption similar to that of Gleason. Therefore Bell's argument in connection with Gleason's work is applicable in the case of Kochen and Specker also. In the case of an angular momentum **J**, for example, it means that hidden variables in general do not assign unique value to $J_{\mathbf{n}}^2$ for each given direction **n**, but, instead, for a given triad of (mutually perpendicular) directions **l, m, n**, they tell which one among J_l^2, J_m^2 and J_n^2 is having a particular value, where J_l, J_m and J_n represent the components of **J**.

Bell's Inequalities and Bell's Theorem

Bell[38] has analysed the EPR thought-experiment from the viewpoint of hidden variables with a view to ascertaining whether hidden-variable theories of the second kind (which satisfy Einstein's locality postulate) could really be consistent with all the predictions of quantum theory. The result is a set of inequalities which bear his name and which establish the *Bell's theorem* which states: *No local hidden-variables theory can reproduce all the results of quantum theory.* These inequalities also permit one to resolve through experiments the question whether quantum theory or a local hidden-variables theory represents the laws of nature. We give below an outline of the derivation of the inequalities:

Consider the two spin $\frac{1}{2}$ particle in a singlet state, moving away in different directions. A Stern-Gerlach magnet measures the component of σ_1 in the direction of a unit vector **a** while another Stern-Gerlach magnet measures the component of σ_2 along the unit vector **b**. $(\sigma_1 \cdot \mathbf{a})$ and $(\sigma_2 \cdot \mathbf{b})$ can have the values ± 1 only. But the statistical correlation between the two measurements is given, according to quantum theory, by

$$\langle (\sigma_1 \cdot \mathbf{a})(\sigma_2 \cdot \mathbf{b}) = -\langle (\sigma_1 \cdot \mathbf{a})(\sigma_1 \cdot \mathbf{b}) \rangle = -(\mathbf{a} \cdot \mathbf{b}), \tag{12.24}$$

where the relationships,

$$\sigma_1 + \sigma_2 = 0$$

and

$$(\sigma \cdot \mathbf{a})(\sigma \cdot \mathbf{b}) = (\mathbf{a} \cdot \mathbf{b}) + i\sigma \cdot (\mathbf{a} \times \mathbf{b}),$$

are used.

Let us suppose that the results of individual measurements, which quantum theory is unable to predict, are determined by a set of parameters (hidden variables) λ. Then the result A of measurement of $(\sigma_1 \cdot \mathbf{a})$ will depend on **a** and λ while the result B of measurement of $(\sigma_2 \cdot \mathbf{b})$ will depend on **b** and λ. Since the

46. Kochen, S. and Specker, E.P. *J. Math. Mech.* **17**, 59 (1967).

result of a measurement has to be an eigenvalue of the corresponding operator and since the eigenvalues of any component of σ are ± 1, we have,

$$A\ (\mathbf{a},\lambda) = \pm 1; B\ (\mathbf{b},\lambda) = \pm 1. \tag{12.25}$$

Einstein's locality postulate requires that A be independent of the choice $\mathbf{b}$ and B be independent of the choice $\mathbf{a}$. If $\rho(\lambda)$ represents the probability distribution of the hidden variables, then the average of the product $(\sigma_1 \cdot \mathbf{a})(\sigma_2 \cdot \mathbf{b})$ in the hidden-variables theory is given by

$$\langle (\sigma_1 \cdot \mathbf{a})(\sigma_2 \cdot \mathbf{b}) \rangle_{h.v} = P\ (\mathbf{a} \cdot \mathbf{b}) \tag{12.26}$$

with

$$P\ (\mathbf{a},\mathbf{b}) = \int A\ (\mathbf{a},\lambda)\, B\ (\mathbf{b},\lambda)\, \rho(\lambda)\, d\ \lambda. \tag{12.26a}$$

If we assume that the measurement results depend also on some external hidden variables (that is, the measuring instruments also have hidden variables) over which averaging has to be done, and if we denote this averaging by a bar, then in place of (12.25) and (12.26a) we would have,[47].

$$|\overline{A}(a,\lambda)| \leq 1\ ;\ |\overline{B}(b,\lambda)| \leq 1 \tag{12.25a}$$

$$P(\mathbf{a},\mathbf{b}) = \int \overline{A}(\mathbf{a},\lambda)\, \overline{B}(\mathbf{b},\lambda)\, \rho\ (\lambda)\, d\lambda. \tag{12.26b}$$

The question is whether the absolute difference between the hidden-variables result (12.26b) and the quantum mechanical result (12.25) could be made arbitrarily small. That is, if we put

$$|P(\mathbf{a},\mathbf{b}) + (\mathbf{a} \cdot \mathbf{b})| \leq \epsilon \tag{12.27}$$

can ϵ be made to vanish? Bell obtains the inequality[48],

$$|P(\mathbf{a}',\mathbf{b}') - P(\mathbf{a},\mathbf{b})| + |P(\mathbf{a}',\mathbf{b}') + P(\mathbf{a},\mathbf{b}')| \leq 2, \tag{12.28}$$

and thence shows that

$$4\epsilon > \sqrt{2} - 1. \tag{12.29}$$

Hence the Bell's theorem.

The Experimental Verdict

The Bell's inequality (12.28) and other similar inequalities[48] for photon polarization correlation measurements, enable one to verify or refute the claims of the local hidden variable theories as against the predictions of quantum theory (the latter will violate the inequalities). A number of experiments have been performed in this connection. Most of them seem to agree with the predictions of quantum theory. In this connection, an interesting thought-experiment, which is

47. Bell, J.S. 'Introduction to the hidden-variable question' in *Foundations of Quantum Mechanics* (Proceedings of the International School of Physics "Enrico Fermi", Course 49, June 29 - July 11, 1979, Vienna), published by Academic Press, New York.
48. See Roy, S.M. *Phys. News.* **11**, No. 1, p. 4 (1980), for other related inequalities.

supposed to reflect the experimental findings of Aspect and coworkers[49] at the University of Paris, has been presented by Mermin[7]. The results are equivalent to the following proposition in relation to the EPR experiment:

> Data gathered on $S^{(1)}$ and data gathered *independently* on $S^{(2)}$, are correlated. This correlation is purely statistical and arises from the fact that $S^{(1)}$ and $S^{(2)}$ have a common (coherent) origin[50].

According to Mermin, this result rules out a deterministic theory like local hidden variable theory since the correlations found are as predicted by quantum theory. However, many believe that the question whether quantum theory or something beyond it is the ultimate physical theory, is far from settled.

---

49. Aspect, A, Grangier, P and Roger, G. *Phys. Rev. Lett.* **47**, 460 (1981); **49**, 91 (1982); Aspect, A, Dalibard, J. and Roger, G. *Phys. Rev. Lett.* **49**, 1804 (1982).
50. It should, however, be remembered that the subject matter of the EPR experiment is not the correlations arising from independent *measurements* done on $S^{(1)}$ *and* $S^{(2)}$. The question posed by the EPR experiment is whether or not quantum mechanics permits the *deduction* of a component of $S^{(2)}$ from *a measurement done on* $S^{(1)}$ *only* (without making a measurement on $S^{(2)}$ itself). As we have seen in Section 12.4, the answer to the question is negative.

APPENDIX A

MATRICES

A1 DEFINITION

A matrix is a two-dimensional array of numbers, in general rectangular, conforming to the following definitions:

$$A \equiv \begin{pmatrix} A_{11} & A_{12} & . & . & . & A_{1n} \\ A_{21} & A_{22} & . & . & . & A_{2n} \\ \vdots & & & & & \\ A_{m1} & A_{m2} & . & . & . & A_{mn} \end{pmatrix} \tag{A.1}$$

The horizontal arrays are called *rows* and the vertical arrays *columns*. A matrix having m rows and n columns, as in (A.1), is called an $(m \times n)$ matrix, or a matrix of *order* $(m \times n)$. A matrix having equal number of rows and columns $(m = n)$, is called a *square matrix*. n is then the *dimension* or order of the matrix. The numbers A_{ij} are called the *elements* of the matrix; the first subscript denoting the row and the second the column in which the element appears.

The elements A_{ii} of a square matrix constitute the *principal diagonal* of the matrix, while the elements themselves are called the *diagonal elements*. The elements A_{ij} for $i \neq j$ are, then, the *off-diagonal elements*. A matrix for which all the non-vanishing elements are diagonal, is called a *diagonal matrix*. Thus, if D is a diagonal matrix, then,

$$D_{ij} = \delta_{ij} D_i. \tag{A.2}$$

A diagonal matrix whose diagonal elements are all unity, is a *unit matrix* and is denoted by I.

Thus,

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 & . & . \\ 0 & 1 & 0 & 0 & . & . \\ 0 & 0 & 1 & 0 & . & . \\ . & . & . & . & . & . \end{pmatrix} \text{ or, } I_{ij} = \delta_{ij}. \tag{A.3}$$

The unit matrix could be of any order.

A *null-* or *zero-matrix* is one whose elements are all zero:

$$0_{ij} = 0. \tag{A.4}$$

A matrix which is part of a larger matrix, is called a *submatrix*.

Example: The matrices

$$P = \begin{pmatrix} 1 & 9 \\ 2 & 8 \end{pmatrix};\ Q = \begin{pmatrix} 3 \\ 4 \end{pmatrix};\ R = (6 \quad 2)$$

and $S = (7)$, are submatrices of order, 2, (2×1), (1×2) and 1, respectively, of the 3rd order matrix

$$A = \begin{pmatrix} 1 & 9 & 3 \\ 2 & 8 & 4 \\ 6 & 2 & 7 \end{pmatrix}$$

A2 MATRIX ALGEBRA

An algebra of matrices can be developed by defining equality, sum, product, etc. of matrices:

Equality: Two matrices A and B are equal if their corresponding elements are equal.

Thus,

$$A = B, \text{ if } A_{ij} = B_{ij}, \text{ for all } i, j. \tag{A.5}$$

Sum:
$$C = A + B = B + A, \text{ if } C_{kl} = A_{kl} + B_{kl}. \tag{A.6}$$

Product: If $C = AB$, then, $C_{ij} = \sum_k A_{ik} B_{kj}$. (A.7)

Thus, if A is $(m \times n)$, B should be $(n \times l)$ while C would be $(m \times l)$. Therefore, existence of AB does not imply existence of BA, so that, in general, $AB \neq BA$. For this reason, AB is called the product of B by A. Matrix algebra is, thus, *non-commutative*. Matrix multiplication is, however,

$$\textit{distributive: } A(B + C) = AB + AC, \tag{A.7a}$$

and
$$\textit{associative: } A(BC) = (AB)C = ABC. \tag{A.7b}$$

It follows from (A.7) and (A.2), that *diagonal matrices commute among themselves.* Also, *a matrix A that commutes with all diagonal matrices is necessarily diagonal.*

Let
$$(AD)_{ij} = (DA)_{ij},$$

i.e.,
$$A_{ij} D_j = D_i A_{ij}, \text{ since } D_{ik} = \delta_{ik} D_i.$$

i.e.,
$$(D_i - D_j) A_{ij} = 0. \tag{A.8}$$

Hence
$$A_{ij} = 0 \text{ if } D_i \neq D_j; \text{ that is, if } i \neq j,$$

or
$$A_{ij} = \delta_{ij} A_i.$$

Multiplication of a matrix by a complex number c, is defined by

$$(Ac)_{ij} = (cA)_{ij} = cA_{ij}. \tag{A.9}$$

Thus, the matrix $C = cI$, is a diagonal matrix whose diagonal elements are all equal to the number c. Such a matrix is called a *constant matrix*:

$$C_{ij} = \delta_{ij} c. \tag{A.10}$$

A constant matrix commutes with all square matrices

$$(CA)_{ij} = \sum_k C_{ik} A_{kj} = cA_{ij};$$

$$(AC)_{ij} = \sum_k A_{ik} C_{kj} = cA_{ij}.$$

Conversely, a *matrix that commutes with all square matrices, is a constant matrix.*

Let A be the matrix that commutes with all square matrices. In particular, A will then commute with all diagonal matrices, so that, according to Eq. (A.8), A is diagonal.

$$A_{ij} = \delta_{ij} A_i$$

Let B be an arbitrary square matrix. Then,

$$AB - BA = 0$$

i.e.,

$$(A_i - A_j) B_{ij} = 0.$$

Since $B_{ij} \neq 0, A_i = A_j$ for all i and j.

This means that all the diagonal elements of A are equal to the same constant.

$$A_{ij} = \delta_{ij} c.$$

Since the unit and the null matrices are also constant matrices, it follows that they commute with all square matrices.

$$AI = IA = A, \tag{A.11}$$

$$AO = OA = O. \tag{A.12}$$

Direct, or *Tensor*, *Product*[1] of two matrices A and B is defined as follows:

If

$$C = A \otimes B,$$

then, the elements of C are given by

$$C_{ik;jl} = A_{ij} B_{kl}, \tag{A.13}$$

Thus, if A is $(m_1 \times n_1)$ and B is $(m_2 \times n_2)$, then C is an $(m_1 m_2 \times n_1 n_2)$ matrix. As an example, let A and B be (2×2) matrices. Then C will be the following (4×4) matrix:

$$A \otimes B \equiv C = \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22} \end{pmatrix}$$

---

1. The names, *Kronecher product* and *outer product*, are also used. It is also possible to define a *direct sum* of two matrices by

$$A \oplus B = \begin{pmatrix} A & O_1 \\ O_2 & B \end{pmatrix},$$

where, if A is $(m \times n)$ and B, $(p \times q)$, then, the null matrix O_1 is $(m \times q)$ and O_2 is $(p \times n)$.

$$= \begin{pmatrix} A_{11}B & A_{12}B \\ A_{21}B & A_{22}B \end{pmatrix} \tag{A.13a}$$

Thus, the tensor product of A and B is a matrix of the type A whose elements are matrices of type B.

Power: The square of a matrix A is defined by

$$A^2 = AA \tag{A.14a}$$

Similarly, $$A^n = AA^{n-1} = \underbrace{A \ldots A}_{n \text{ factors}} \tag{A.14b}$$

Also, $$A^0 = I. \tag{A.14c}$$

Obviously, only square matrices have powers.

Function: Function of a matrix can be defined by combining the operations of addition, multiplication and power. Thus,

$$f(A) \equiv aA^2 + bA + cI, \tag{A.15}$$

is a function of the matrix A, whereas the equation $f(A) = 0$, is called a *matrix equation.*

Inverse: B is said to be inverse of the matrix A, if

$$BA = AB = I. \tag{A.16}$$

B is, then, written as A^{-1}. If A is $(m \times n)$, then A^{-1} has to be $(n \times m)$. But (AA^{-1}) and $(A^{-1}A)$ should both be square matrices of the same order, according to Eq. (A.16). Therefore, $m = n$, so that, *only square matrices have inverses.* Every square matrix, however, need not have an inverse. If it has, it is said to be a *non-singular* matrix, whereas if it does not have, it is a *singular matrix.*

The inverse of a product of matrices is the product of the inverse of matrices in the reverse order:

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1}, \tag{A.17}$$

Transpose: The matrix obtained by interchanging the rows and columns of the matrix A, is called the transpose of A, and is denoted by $\tilde{A}$.
Thus,

$$\tilde{A}_{ij} = A_{ji}. \tag{A.18}$$

The transpose of a product of matrices is the product of the transpose of the matrices in the reverse order:

$$(\tilde{ABC}) = \tilde{C}\tilde{B}\tilde{A}. \tag{A.19}$$

Complex Conjugate: The complex conjugate A^* of the matrix A is defined by,

$$(A^*)_{ij} = (A_{ij})^* \tag{A.20}$$

Also, $$(ABC)^* = A^*B^*C^*. \tag{A.21}$$

Adjoint (Hermitian Conjugate): The adjoint $A^\dagger$ of the matrix A is given by

$$A^\dagger = (\tilde{A})^* = (\tilde{A^*}),$$

or,
$$(A^\dagger)_{ij} = (A_{ji})^*. \tag{A.22}$$

From (A.19) and (A.21), we have,

$$(ABC)^\dagger = C^\dagger B^\dagger A^\dagger. \tag{A.23}$$

A3 IMPORTANT SCALAR NUMBERS ASSOCIATED WITH A SQUARE MATRIX

Trace: This is the sum of the diagonal elements and is denoted by Tr(A). Thus,

$$\mathrm{Tr}(A) = \sum_i A_{ii}. \tag{A.24}$$

Using Eqs. (A.7) and (A.24), we have,

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA). \tag{A.24a}$$

Also, from Eq. (A.13a), we get,

$$\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A) \cdot \mathrm{Tr}(B). \tag{A.24b}$$

Determinant: This is defined for a square matrix A by,

$$\det A \equiv |A| = \begin{vmatrix} A_{11}A_{12} & \cdots & A_{1N} \\ A_{21}A_{22} & \cdots & A_{2N} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ A_{N1}A_{N2} & \cdots & A_{NN} \end{vmatrix}$$

$$= \sum_{i_1, i_2 \ldots} \pm A_{1i_1} A_{2i_2} \ldots A_{Ni_N}; \tag{A.25}$$

where, $i_1 i_2 \ldots i_N$, is one of the $N!$ permutations of the numbers $1, 2, \ldots, N$. The + sign is to be chosen when the permutation is even (or, cyclic) and the – sign when the permutation is odd.

Example:

$$\text{Let } A \equiv \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \tag{A.26}$$

Then,

$$\begin{aligned} \det A = {} & A_{11}(A_{22}A_{33} - A_{23}A_{32}) \\ & - A_{12}(A_{21}A_{33} - A_{23}A_{31}) \\ & + A_{13}(A_{21}A_{32} - A_{22}A_{31}). \end{aligned} \tag{A.27}$$

Determinant of a product of matrices is equal to the product of the determinants:

$$|ABC| = |A| \cdot |B| \cdot |C|. \tag{A.28}$$

Obviously, $\det(AB) = \det(BA)$. (A.29)

The *order* of $\det A$, is the order of the matrix A.

Minor: The minor M_{ij} of the element A_{ij} of the matrix A, is the determinant of the matrix obtained by removing the ith row and the jth column (that is, the row and the column in which the element A_{ij} occurs):

$$M_{ij} = \begin{vmatrix} A_{11} & A_{12} & \cdots & A_{1j-1} & A_{1j+1} & .. & A_{1N} \\ . & . & . & . & . & . & . \\ A_{i-11} & A_{i-12} & \cdots & A_{i-1j-1} & A_{i-1j+1} & .. & A_{i-1N} \\ A_{i+11} & A_{i+i2} & \cdots & A_{i+1j-1} & A_{i+1j+1} & .. & A_{i+1N} \\ . & . & . & . & . & . & . \\ A_{N1} & A_{N2} & \cdots & A_{Nj-1} & A_{Nj+1} & .. & A_{NN} \end{vmatrix} \tag{A.30}$$

In the example (A.26), we have,

$$M_{12} = \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} = A_{21} A_{33} - A_{23} A_{31}$$

Cofactor: Cofactor a_{ij} of the element A_{ij} is defined as,

$$a_{ij} = (-1)^{i+j} M_{ij}. \tag{A.31}$$

a_{ij} could be considered as the ijth element of an $(N \times N)$ matrix a.

Thus, in the above example,

$$a_{12} = -M_{12} = A_{23} A_{31} - A_{21} A_{33}.$$

The determinant of a matrix can be expressed in terms of the cofactors of its elements:

$$\det A = \sum_{j=1}^{N} A_{ij} a_{ij} = \sum_{i=1}^{N} A_{ij} a_{ij}. \tag{A.32}$$

This formula is known as the *Laplace development.*

For the matrix A given by (A.26), we get,

$$\det A = A_{11} a_{11} + A_{12} a_{12} + A_{13} a_{13}$$

$$= A_{11} \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{12} \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} + A_{13} \begin{vmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{vmatrix}$$

$$= A_{11} (A_{22} A_{33} - A_{23} A_{32}) - A_{12} (A_{21} A_{33} - A_{23} A_{31}) + A_{13} (A_{21} A_{32} - A_{22} A_{31}),$$

which is the same as (A.27).

The following properties of the determinant of a matrix follows from the definition (A.32):

(i) $\det \tilde{A} = \det A$; $\det A^{\dagger} = (\det A)^{*}$,

(ii) $\det A_{or} = \det A_{oc} = -\det A$,

where, A_{or} is the matrix obtained from A by an odd permutation of its rows, while A_{oc} is obtained from A by an odd permutation of its columns. It follows, from this, that if any two rows or any two columns of A are identical, then $\det A = 0$.

(iii) If B is a matrix which is obtained from A by multiplying all the elements of any *one* row or any one column of A by the number c, then, $\det B = c \det A$.

(iv) If each element in any row or any column of A is written as a sum of two numbers, then $\det A$ can be written as the sum of two determinants:

Thus, if

$$A = \begin{pmatrix} A_{11} & A_{12}^{(1)} + A_{12}^{(2)} & A_{13} \\ A_{21} & A_{22}^{(1)} + A_{22}^{(2)} & A_{23} \\ A_{31} & A_{32}^{(1)} + A_{32}^{(2)} & A_{33} \end{pmatrix}$$

then, $\det A = \det A^{(1)} + \det A^{(2)}$, where

$$A^{(k)} = \begin{pmatrix} A_{11} & A_{12}^{(k)} & A_{13} \\ A_{21} & A_{22}^{(k)} & A_{23} \\ A_{31} & A_{32}^{(k)} & A_{33} \end{pmatrix}$$

From properties (ii), (iii) and (iv), it follows that if we add any multiple of a row (column) of A to any *other* row (column), then the determinant of the resulting matrix is the same as $\det A$.

The inverse also can be expressed in terms of the cofactors and the determinant of the matrix:

Eq. (A.32) can be written as

$$|A| = \sum_{j=1}^{N} A_{ij}\tilde{a}_{ji} = D_{ii}, \tag{A.33}$$

where, $\tilde{a}$ is the transpose of the matrix a whose elements are the cofactors of A. Thus,

$$D = A\tilde{a} = |A| I = I |A|, \tag{A.34}$$

is a constant matrix whose diagonal elements are all equal to $|A|$. If A has an inverse, then,

$$I = AA^{-1},$$

so that Eq. (A.34) becomes,

$$A\tilde{a} = A(A^{-1}|A|),$$

or,

$$A^{-1} = \frac{\tilde{a}}{|A|}. \tag{A.35}$$

Hence, the condition for the existence of an inverse is that the determinant of the matrix be not equal to zero. This condition is both necessary and sufficient.

Rank: The rank $\rho(A)$ of the matrix A is defined as the order of the largest non-singular (sub) matrix contained within A. Thus, if A is non-singular,

$$\rho(A) = n, \tag{A.36a}$$

where n is the order of the matrix.

If A is singular,

$$\rho(A) < n. \tag{A.36b}$$

Example 1:

$$A = \begin{pmatrix} 1 & 3 & 4 \\ 2 & 4 & 5 \\ 3 & 1 & 2 \end{pmatrix}.$$

$|A| \neq 0$. Therefore, $\rho(A) = 3$.

Example 2:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \end{pmatrix}$$

$|A| = 0$. Also, all submatrices of order 2 are singular. Hence $\rho(A) = 1$.

Example 3:

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 1 & -1 \end{pmatrix}$$

$$\rho(A) = 2, \text{ since} \begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} \neq 0.$$

The rank of a product of two matrices is less than or equal to the rank of either.

$$\rho(AB) \leq \rho(A), \tag{A.36c}$$

where $\rho(A) \leq \rho(B)$.

A4 SPECIAL MATRICES

Hermitian: A is Hermitian or *self-adjoint* if

$$A^\dagger = A, \text{ or, } A^*_{ji} = A_{ij} \tag{A.37a}$$

and *antiHermitian*, if

$$A^\dagger = -A; \text{ or, } A^*_{ji} = -A_{ij} \tag{A.37b}$$

Unitary: U is called a Unitary matrix, if

$$U^\dagger = U^{-1}, \tag{A.38a}$$

or,

$$U^\dagger U = UU^\dagger = 1. \tag{A.38b}$$

Orthogonal: If

$$A = \tilde{A}^{-1}, \text{ or } A\tilde{A} = \tilde{A}A = I, \tag{A.39}$$

A is said to be orthogonal.

Symmetric: When

$$\tilde{A} = A, \text{ or } A_{ij} = A_{ji}, \tag{A.40a}$$

A is symmetric, and if

$$\tilde{A} = -A, \text{ or, } A_{ij} = -A_{ji}, \tag{A.40b}$$

A is *antisymmetric.*

Real: A matrix is said to be real, if

$$A^* = A, \text{ or } A^*_{ij} = A_{ij}, \tag{A.41a}$$

and *imaginary* if

$$A^* = -A; \text{ i.e., } A^*_{ji} = -A_{ij}. \tag{A.41b}$$

Normal: A matrix N that commutes with its adjoint $N^\dagger$, is said to be a Normal matrix:

$$NN^\dagger = N^\dagger N. \tag{A.42}$$

Obviously, Hermitian, Unitary and real-symmetric matrices are Normal.

Column and Row Matrices: A matrix X with *only one* column is called a *column-matrix* or a *ket-vector*. Similarly, a matrix Y with only one row, is a *row-matrix*, or a *bra-vector*.[2] The name 'vector' in the case of a column (or row) matrix has its origin in the fact that the elements of such a matrix could be regarded as the components of a vector. For example, let x, y, z be the components of the position vector $\mathbf{r}$:

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

This vector could be represented by the column matrix,

$$r \equiv \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

The scalar product of two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ is defined as

$$\mathbf{r}_1 \cdot \mathbf{r}_2 = x_1x_2 + y_1y_2 + z_1z_2,$$

which is the product (see Eq. (A.7)) of the row-matrix

$$r_1^\dagger = \tilde{r}_1 = (x_1\ y_1\ z_1) \equiv x_1\ \ y_1\ \ z_1$$

and the column matrix,

$$r_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}.$$

---

2. These are also sometimes referred to as column-vector and row-vector, respectively.

Thus, both the column- and the row-matrices represent vectors. The names ket and bra-vectors, are due to P.A.M. Dirac (See Section 2.3).

The above concept of vectors could be generalized to the case of a complex space with more than three dimensions. Vectors in such spaces would be represented by column or rowmatrices with complex elements.

Whereas the product of a column-matrix by a row-matrix is a scalar, the product of a row-matrix by a column matrix is a square matrix:

Let $$X^\dagger = x_1^* x_2^* \ldots x_N^*$$

and $$Y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_N \end{pmatrix}$$

Then, $$X^\dagger Y = \sum_{i=1}^{N} x_i^* y_i. \tag{A.43}$$

But, $$YX^\dagger = \begin{pmatrix} y_1x_1^* & y_1x_2^* & \ldots & y_1x_N^* \\ y_2x_1^* & y_2x_2^* & \ldots & y_2x_N^* \\ y_Nx_1^* & y_Nx_2^* & \ldots & y_Nx_N^* \end{pmatrix} \tag{A.44}$$

A5 MATRIX TRANSFORMATIONS

A *similarity transformation* of a square matrix A by a nonsingular matrix S, is defined by

$$A \rightarrow A' = S^{-1}AS. \tag{A.45}$$

If S is a unitary matrix, the transformation is called *unitary*. If A is a nondiagonal matrix but A' is diagonal, then, A is said to be *diagonalized* by the transformation, or *by the matrix S*. If two matrices *A and B commute, then, they can be diagonalized by the same matrix S*:

$$AB = BA, \text{ (given)}.$$

Let $A' = S^{-1}AS$, be diagonal. We have to show that $B' = S^{-1}BS$, is also diagonal.

We have, $$S^{-1}(AB - BA)S = S^{-1}OS = 0. \tag{A.46}$$

i.e., $$S^{-1}ASS^{-1}BS - S^{-1}BSS^{-1}AS = 0,$$

or, $$A'B' - B'A' = 0. \tag{A.47}$$

i.e., $$(A'_i - A'_j)B'_{ij} = 0. \tag{A.48}$$

Hence, $$B'_{ij} = \delta_{ij}B'_i. \tag{A.49}$$

Conversely, if *A and B are diagonalized by the same matrix S, then, AB = BA*:

We are given, $A'B' - B'A' = 0.$

That is (see Eq. (A.46)),

$$S^{-1}(AB - BA)S = 0.$$

Multiplying from the left by S and from the right by S^{-1}, we get,

$$AB - BA = SOS^{-1} = 0. \tag{A.50}$$

Obviously (see Eqs. (A.24a) and (A.29)), the trace and the determinant of a matrix are invariant under a similarity transformation:

$$\text{Tr}\,(A') = \text{Tr}\,(S^{-1}AS) = \text{Tr}\,(S^{-1}SA) = \text{Tr}\,(A) \tag{A.51}$$

$$\det A' = \det (S^{-1}AS) = \det (S^{-1}S) \cdot \det A = \det A. \tag{A.52}$$

The rank of a matrix is another quantity which is invariant under similarity transformation. Also, a matrix equation is unaffected if every matrix in the equation is subjected to the same similarity transformation.

A unitary transformation (but not a similarity transformation by a non-unitary matrix) preserves the Hermitian, or Unitary, character of a matrix, since

$$A'^{\dagger} = (U^{-1}AU)^{\dagger} = U^{\dagger}A(U^{-1}) = U^{-1}AU = A', \text{ when } A^{\dagger} = A,$$

and $$A'^{\dagger} = U^{\dagger}A^{-1}U = U^{-1}A^{-1}U = (A')^{-1} \qquad \text{when } A^{\dagger} = A^{-1}.$$

A6 SOLUTION OF LINEAR ALGEBRAIC EQUATIONS

A set of linear algebraic equation in n variables, $x_1, x_2, \ldots, x_n$, is given by

$$A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n = y_1$$

$$A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n = y_2$$

$$\vdots$$

$$A_{n1}x_1 + A_{n2}x_2 + \ldots + A_{nn}x_n = y_n.$$

or, using matrices,

$$AX = Y \tag{A.53}$$

where, A is the *coefficient matrix*, given by,

$$A = \begin{pmatrix} A_{11} & \ldots & A_{1n} \\ \cdot & & \\ \cdot & & \\ A_{n1} & \ldots & A_{nn} \end{pmatrix}, \tag{A.53a}$$

and X and Y are column matrices (vectors):

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}; \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}. \tag{A.53b}$$

The problem is to determine X when A and Y are given[3].

Multiplying both sides of (A.53) with A^{-1}, we get,

$$X = A^{-1} Y = \frac{\tilde{a}\, Y}{|A|}, \tag{A.54}$$

where, we have substituted for A^{-1} from Eq. (A.35).
That is,

$$x_j \equiv (X)_{j1} = \frac{1}{|A|} \sum_{i=1}^{n} \left(\tilde{a}\right)_{ji} y_i$$

$$= \frac{1}{|A|} \sum_{i=1}^{n} y_i\, a_{ij} = \frac{|A_y^j|}{|A|}, \tag{A.55}$$

where, [see Eq. (A.33)], A_y^j is the matrix obtained from A by replacing its jth column with Y.

As an example of the application of formula (A.55), let us consider the set of equations,

$$x_1 - 2x_2 + 3x_3 = 2$$

$$2x_1 - 3x_3 = 3$$

$$x_1 + x_2 + x_3 = 6.$$

Here,

$$A = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 0 & -3 \\ 1 & 1 & 1 \end{pmatrix}; \quad |A| = 19,$$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}; \quad Y = \begin{pmatrix} 2 \\ 3 \\ 6 \end{pmatrix}.$$

$$\therefore\ x_1 = \frac{1}{19} \begin{vmatrix} 2 & -2 & 3 \\ 3 & 0 & -3 \\ 6 & 1 & 1 \end{vmatrix} = \frac{57}{19} = 3 = \frac{|A_y^1|}{|A|};$$

$$x_2 = \frac{1}{19} \begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & -3 \\ 1 & 6 & 1 \end{vmatrix} = \frac{38}{19} = 2 = \frac{|A_y^2|}{|A|};$$

---

3. The equation (A.53) will have solutions only if

$$\text{rank } (A) = \text{rank } (A, Y),$$

where (A, Y) is the *extended* or *augmented* matrix, obtained by attaching Y to A as the $(n+1)$th column.

$$x_3 = \frac{1}{19}\begin{vmatrix} 1 & -2 & 2 \\ 2 & 0 & 3 \\ 1 & 1 & 6 \end{vmatrix} = \frac{19}{19} = 1 = \frac{|A_y^3|}{|A|}.$$

When,

$$AX = 0, \tag{A.56}$$

we have, $|A_y^j| = 0$, for all j; so that, according to (A.55), a non-trivial solution (one for which $X \neq 0$) for (A.56) exists only when

$$|A| = 0. \tag{A.57}$$

A7 EIGENVALUES AND EIGENVECTORS

Consider a square matrix A. The matrix $K = (A - \alpha I)$, where α is a scalar number [defined by Eq. (A.58) below] and I is the unit matrix of the same order as A, is called the *characteristic matrix* of A. The determinant of K is the *characteristic function* (or, characteristic polynomial) and

$$\det K \equiv |A - \alpha I| = 0, \tag{A.58}$$

is called the *characteristic*, or *secular equation* of A. If the order of A is n, then, the secular equation is of degree n in α. The n roots of the equation are called the *eigenvalues* of A.

Now, if A is a diagonal matrix, $A_{ij} = \delta_{ij}A_i$, and the secular equation is,

$$|A - \alpha I| = \prod_{i=1}^{n} (A_i - \alpha) = 0. \tag{A.59}$$

Hence, the roots are

$$\alpha_i = A_i, i = 1, 2, \ldots, n. \tag{A.60}$$

Thus, the eigenvalues of a diagonal matrix are the diagonal elements of the matrix.

Suppose X is a column matrix of order $n \times 1$.

Then,

$$AX = Y, \tag{A.61}$$

where, Y is also of order $n \times 1$. Y is then called the *transform* of the vector X by the matrix A. There might be vectors X for which $Y = \alpha X$. That is,

$$AX = \alpha X, \tag{A.62}$$

or

$$(A - \alpha I)X \equiv KX = 0. \tag{A.63}$$

Eq (A.63) represents a set of n linear, homogeneous, algebraic equations. For a non-trivial solution, the condition is that [see Eqs. (A.56) and (A.57)]

$$\det K \equiv |A - \alpha I| = 0.$$

But, this is the secular equation of A. Eq. (A.58)), so that the values of α for which Eq. (A.62) is satisfied, are the eigenvalues of A. Eq. (A.62) is, for this reason, called the *eigenvalue equation* of A. Denoting by α_k the different values of α, we have,

$$AX_k = \alpha_k X_k, \; k = 1, 2, \ldots, n. \tag{A.62a}$$

X_k is the *eigenvector*[4] of A belonging to the eigenvalue α_k.

The eigenvalues are invariant under a similarity transformation. For, let

$$A' = S^{-1}AS.$$

Then, the secular equation for A' is,

$$|A' - \alpha' I| = 0 \tag{A.64}$$

That is,

$$|S^{-1}AS - \alpha' I| = |S^{-1}AS - S^{-1}\alpha' IS| = 0$$

i.e.,

$$|S^{-1}(A - \alpha' I)S| = 0.$$

Using Eq. (A.28), we get,

$$|S^{-1}| \cdot |A - \alpha' I| \cdot |S| = 0.$$

Since S is non-singular, $|S^{-1}|$ and $|S|$ are not zero, so that

$$|A - \alpha' I| = 0. \tag{A.65}$$

This equation is the secular equation (A.58) for A. Therefore, the roots α_k' are the same as the eigenvalues α_k of A.

Since the eigenvalues of a diagonal matrix are its diagonal elements, and since the eigenvalues are invariant under similarity transformations, one way of finding the eigenvalues would be by diagonalising the matrix through a similarity transformation. Another method, of course, would be by solving the characteristic equation.

Now, each of the vectors X_k in Eq. (A.62a) is a column matrix:

$$X_k = \begin{pmatrix} x_{1k} \\ x_{2k} \\ \cdot \\ \cdot \\ \cdot \\ x_{nk} \end{pmatrix}$$

Define the square matrices S and A' by,

$$S = (X_1 X_2 \ldots X_n) = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \cdot & & & \\ \cdot & & & \\ x_{n1} & x_{n2} & \ldots & x_{nn} \end{pmatrix} \tag{A.66}$$

---

4. If X is an eigenvector of A, then cX, where c is a scalar number, is also an eigenvector. However, X and cX are not counted as separate vectors. To avoid the arbitrariness in the selection of the eigenvectors, the eigenvectors are *normalized*; that is, X is so chosen that

$$X^*X = 1.$$

$$A' = \begin{pmatrix} \alpha_1 & 0 & 0.....0 \\ 0 & \alpha_2 & 0.....0 \\ . & . & .... \\ . & . & .... \\ 0 & 0 & 0....\alpha_n \end{pmatrix} \tag{A.67}$$

Eq. (A.62a) could be written as,

$$(AX_1\, AX_2 \ldots AX_n) = (X_1\alpha_1\, X_2\alpha_2 \ldots X_n\alpha_n),$$

or,
$$AS = SA'. \tag{A.68}$$

Multiplying from the left by S^{-1}, we get,

$$A' = S^{-1}AS. \tag{A.69}$$

Thus, the matrix S that diagonalizes A, has the eigenvectors of A as its columns. The problem of finding the eigenvalues and eigenvectors of A is reduced to the problem of finding the matrix S that diagonalizes A.

Now, the secular equation (A.58) can be written as,

$$\alpha^n + c_1\alpha^{n-1} + \ldots + c_n = 0. \tag{A.70}$$

Define,
$$s_n = \text{Tr}\,(A^n),\ n \text{ a positive integer.} \tag{A.71}$$

Then, it has been shown that the coefficients in (A.70) are given by

$$c_r = -\frac{1}{r}\sum_{k=1}^{r} c_{r-k}\, s_k, \tag{A.72}$$

where $c_0 = 1$.

Thus,
$$c_1 = -s_1;\ c_2 = -\tfrac{1}{2}(c_1 s_1 + s_2), \text{ etc.}$$

From the theory of algebraic equations, we have,

$$\alpha_1 + \alpha_2 + \ldots + \alpha_n = \sum_{k=1}^{n} \alpha_k = -c_1 = s_1 = \text{Tr}\,(A) \tag{A.73a}$$

$$\alpha_1\alpha_2 \ldots \alpha_n = (-1)^n c_n = \det A. \tag{A.73b}$$

Eq. (A.73b) follows from the invariance of the determinant as well as the eigenvalues under similarity transformations [Eqs. (A.52), (A.64) and (A.65)] and the fact that the determinant of a diagonal matrix is the product of its diagonal elements which are also its eigenvalues [Eq. (A.60)].

The invariance of the trace and the determinant of a square matrix under similarity transformation, thus, follows from the invariance of the eigenvalues.

Multiplying Eq. (A.70) by an eigenvector X_k of A, and using the fact that,

$$A^r X_k = (\alpha_k)^r X_k, \tag{A.74}$$

we get,

$$A^n + c_1 A^{n-1} + \ldots + c_n I = 0. \tag{A.75}$$

That is, *a square matrix satisfies its own characteristic equation in the matrix form.* This is known as the *Cayley-Hamilton Theorem*[5]. The theorem is useful in finding the inverse of a matrix. For, multiplying both sides of Eq. (A.75) by A^{-1}, we get,

$$A^{-1} = -\frac{1}{c_n}[A^{n-1} + c_1 A^{n-2} + \ldots + c_{n-1} I], \tag{A.76}$$

where, the coefficients c_r are given by Eq. (A.72).

Example: Consider the matrix

$$A = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{A.77}$$

The secular equation is

$$|A - \alpha I| = \begin{vmatrix} -\alpha & -i \\ i & -\alpha \end{vmatrix} = 0,$$

i.e., $$\alpha^2 - 1 = 0.$$

Hence, the eigenvalues are

$$\alpha_1 = 1 \text{ and } \alpha_2 = -1.$$

Substituting these values in the eigenvalue equation

$$AX_k = \alpha_k X_k,$$

and, normalizing X_k, that is, by putting

$$X_k^{\dagger} X_k = 1,$$

we determine

$$X_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix}; \quad X_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-i}{\sqrt{2}} \end{pmatrix}.$$

Thus, the matrix S that diagonalizes A is given by

$$S = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{-i}{\sqrt{2}} \end{pmatrix}. \tag{A.78}$$

In order to show that $S^{-1}AS$ is indeed the matrix

$$A' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

we should first find S^{-1}. According to Eq. (A.76),

$$S^{-1} = -\frac{1}{c_2}(S + c_1 I),$$

---

5. For a derivation of the theorem without using the eigenvector, see Joshi, A.W. *Matrices and Tensors in Physics* (Wiley Eastern, New Delhi 1975), Section 10.

where (see Eqs. (A.71, 72)),

$$c_1 = -Tr(S) = \frac{1}{\sqrt{2}}(i-1);$$

$$c_2 = -\frac{1}{2}(c_1 s_1 + s_2) = -\frac{1}{2}[c_1 Tr(S) + Tr(S^2)]$$

$$= -\frac{1}{2}[-\{Tr(S)\}^2 + Tr(S^2)]$$

$$\{Tr(S)\}^2 = \left\{\frac{1}{\sqrt{2}}(1-i)\right\}^2 = \frac{1}{2}(1-1-2i) = -i$$

$$S^2 = \frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1+i & i-1 \end{pmatrix}$$

$$\therefore\ Tr(S^2) = i$$

Thus,

$$c_2 = -\frac{1}{2}(i+i) = -i.$$

and

$$S^{-1} = -i\left[S + \frac{1}{\sqrt{2}}(i-1)I\right] = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \end{pmatrix} = S^\dagger.$$

Thus, S is unitary, as it should be since A is Hermitian.

The following properties relating to the eigenvalues and eigenvectors of Hermitian and Unitary matrices, are easily proved:

(i) The eigenvalues of a Hermitian matrix are real, whereas the eigenvalues of a Unitary matrix are complex numbers of absolute value unity.

(ii) The eigenvectors belonging to different eigenvalues are orthogonal.

(iii) There are n linearly independent[6] (and, hence, orthonormal) eigenvectors, where n is the order of the matrix[7].

A8 DIAGONALIZABILITY OF A MATRIX

From Eq. (A.69), we see that the condition for the diagonalizability of a square matrix A, is that the matrix S whose columns are the normalized eigenvectors of A, be non-singular. And this would be so when A has n (where n is the order of A) linearly independent eigenvectors. Now, even though every square-matrix of

6. See Eq. (2.17), for the definition of 'linear independence'.
7. See reference given in footnote 5, for proof.

order n has exactly n eigenvalues[8] as implied by the secular equation (A.58) (which is of degree n in α), it need not have n linearly independent eigenvectors. For example, consider the matrix A given by[9]

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}. \tag{A.79}$$

The secular equation is

$$|A - \alpha I| = (1 - \alpha)(\alpha - 2)^2 = 0,$$

so that, the eigenvalues are,

$$\alpha_1 = 1;\ \alpha_2 = \alpha_3 = 2.$$

Substituting these eigenvalues in the eigenvalue equation (A.62a), we get the eigenvectors. It turns out that there are only two linearly independent eigenvectors[10]. These are, corresponding to α_1.

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix},$$

corresponding to α_2 and α_3,

$$X_2 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

Note that X_1 and X_2 are not orthogonal to each other, but are linearly independent. We can find a third vector X_3 which is linearly independent of X_1 and X_2. For example,

---

8. These eigenvalues need not be all different. If α_k is an eigenvalue which occurs k times, then k is called the *multiplicity* of the eigenvalue α_k.
9. This example is taken from the book by Joshi, A.W. (see footnote 5).
10. This is not due to the degeneracy of the roots, but is a consequence of the nature of the matrix. For example, the diagonal matrix

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

has also the roots $\alpha_1 = 1, \alpha_2 = \alpha_3 = 2$. But there are two linearly independent eigenvectors corresponding to the root 2, namely,

$$X_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } X_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

whereas, corresponding to $\alpha_1 = 1$, we have

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

$$X_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

But X_3 is not an eigenvector of A belonging to eigenvalue 2. Also, the matrix $S \equiv (X_1\, X_2\, X_3)$ does not diagonalize A. In fact,

$$S^{-1}AS = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{pmatrix};$$

which is a *triangular* matrix.

Here, we have used,

$$S^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix},$$

which can be obtained using Eq. (A.76).

Thus, every square matrix is not diagonalizable. However, as stated earlier, every finite-order Hermitian or Unitary matrix has as many orthonormal eigenvectors as the order of the matrix. It follows that every Hermitian or Unitary matrix is diagonalizable. That is, *corresponding to every Hermitian (unitary) matrix, there is a unitary matrix that diagonalizes it.*

A9 BILINEAR, QUADRATIC AND HERMITIAN FORMS

If X and Y are column matrices of order $(n \times 1)$, and A is a square matrix of order n, then, the number,

$$X^{\dagger}AY = \sum_{i,j=1}^{n} A_{ij} x_i^{*} y_j; \tag{A.80}$$

is called a *bilinear form* in the $2n$ variables $x_i (i = 1, 2, \ldots, n)$ and $y_i (1, 2, \ldots, n)$. Similarly, the number,

$$X^{\dagger}AX = \sum_{i,j=1}^{n} A_{ij} x_i^{*} x_j, \tag{A.81}$$

is called a *quadratic form.*

If in Eq. (A.81), the matrix A is Hermitian, then, the expression (A.81) is called a *Hermitian form.* A Hermitian form is always real. For, it is clear that a Hermitian form is Hermitian:

$$(X^{\dagger}AX)^{\dagger} = X^{\dagger}A^{\dagger}X = X^{\dagger}AX. \tag{A.82}$$

But, since the Hermitian form is a number,

$$(X^{\dagger}AX)^{\dagger} \equiv (X^{\dagger}AX)^{*}. \tag{A.83}$$

From (A.82) and (A.83), it follows that $X^{\dagger}AX$ is real.

A10 INFINITE MATRICES

The discussion, thus far, has been concerned with *finite matrices*; that is, matrices with finite number of rows and columns. A finite matrix is invariably *discrete*; that is, its elements are labelled by discrete indices i, j where i and j are positive integers. We will now extend the discussion to the case of *infinite matrices*. These are matrices of infinite order, hence having infinite number of rows or columns or both. If the number of rows and columns are *denumerably*[11] infinite, then the matrix is discrete, and its rows and columns are labelled by integers, $1 \leq i, j \leq \infty$. If the rows and columns are labelled by continuous variables x and y, with $a \leq x \leq b$ and $c \leq y \leq d$, then, the number of rows and columns are *nondenumerably* infinite, and the matrix is said to be *continuous*. We can also have a mixed matrix whose row is labelled by discrete numbers and columns by a continuous variable; or *vice versa*.

We will denote the elements of a discrete matrix by A_{ij}, of a continuous matrix by A_{xy} or $A(x, y)$ and of a mixed matrix by A_{ix} or A_{xi}. Also, we will use $A_{\mu\nu} \equiv A(\mu, \nu)$ to denote the elements of a general infinite matrix (μ and ν could be discrete, continuous or partly discrete and partly continuous).

Many of the definitions and operations relating to finite matrices are also valid for infinite matrices. We list below some of the important differences:

1. Two matrices are equal only if their rows and their columns are labelled by the same scheme. That is, if one is discrete, the other is also discrete; if one is continuous, say with row labelled by x and column by y, where $a \leq x \leq b$ and $c \leq y \leq d$, then the other matrix is also continuous with elements A_{xy}, where x and y vary in the interval $a - b$ and $c - d$, respectively.
2. Sum is defined only for those matrices which have their rows labelled by the same scheme and their columns labelled by the same scheme:

   If $$C = A + B \tag{A.84}$$

   then, either $$C_{ij} = A_{ij} + B_{ij}, 1 \leq i, j \leq \infty, \tag{A.84a}$$

   or, $$C(x, y) = A(x, y) + B(x, y); \tag{A.84b}$$

   where, $$a \leq x \leq b; c \leq y \leq d.$$

3. The product AB exists provided the columns of A and the rows of B are labelled by the same scheme and provided, further, that the sum or the integral involved converges. That is, if

   $$C = AB, \tag{A.85}$$

   then, $$C_{\mu\nu} = \sum_{\sigma} A_{\mu\sigma} B_{\sigma\nu}, \tag{A.85a}$$

   where, the summation should be replaced by integration whenever σ is continuous. The product exists provided the R.H.S. of (A.85a) is finite for all values of μ and ν in their allowed range.

---

11. See footnote 1, Chapter 2, for a definition of this term.

Thus, if A and B are discrete, Eq. (A.85a) reads,

$$C_{ij} = \sum_{k=1}^{\infty} A_{ik} B_{kj}, \quad 1 \leq i, j \leq \infty, \tag{A.85b}$$

while if A and B are continuous, we have,

$$C(x, y) = \int_r^s A(x, q) B(q, y) dq. \tag{A.85c}$$

4. A matrix is square if its rows and columns are labelled by the same system of indices (scheme). Thus, the matrix with elements A_{ij}, where $1 \leq i, j \leq \infty$, is square, whereas the continuous matrix with elements A_{xy} is square only if x and y vary in the same interval.

5. For a diagonal matrix, the elements are given by,

$$D_{\mu\nu} = D_{\mu} \delta(\mu, \nu), \tag{A.86}$$

where, $\delta(\mu, \nu) = \delta_{\mu\nu}$,

if μ and ν are discrete, and

$$\delta(\mu, \nu) = \delta(\mu - \nu),$$

if μ and ν are continuous, $\delta(\mu - \nu)$ being the Dirac delta function (see Appendix D). Thus, if D^1 and D^2 are two continuous diagonal matrices, we have,

$$(D^1 D^2)_{\mu\nu} = \int D^1_{\mu} D^2_{\sigma} \delta(\mu - \sigma) \delta(\sigma - \nu) d\sigma$$

$$= D^1_{\mu} D^2_{\nu} \delta(\mu - \nu) = (D^2 D^1)_{\mu\nu}.$$

Hence, infinite diagonal matrices also commute among themselves.

From Eq. (A.86), it follows that the elements of the continuous unit matrix are,

$$I_{\mu\nu} = \delta(\mu - \nu), \tag{A.87}$$

so that, the diagonal elements are not equal to unity, although the off diagonal elements are zero.

6. Determinant of an infinite matrix is not defined. This has the following consequence:

For finite matrices, the relation

$$AB = I, \tag{A.88a}$$

implies that B is the inverter of A, so that,

$$BA = I. \tag{A.88b}$$

For, taking the determinant of (A.88a), we have,

$$\det A \cdot \det B = 1,$$

so that, $\det A \neq 0$, which requires that A has an inverse A^{-1} such that $AA^{-1} = I$.

Since infinite matrices have no determinants, this argument cannot be used to establish the existence of A^{-1} from Eq. (A.88a). Thus, both (A.88a) and (A.88b) have to be satisfied in order that B be the inverse of A.

It also follows that, unlike in the case of finite matrices, A need not be a square matrix to have an inverse. When A is not square, the rows and columns of the unit matrix in (A.88a) would be labelled differently from those of the unit matrix in (A.88b).

The preceding remarks apply to unitary matrices also. Both the conditions,

$$UU^\dagger = I_1 \text{ and } U^\dagger U = I_2, \tag{A.89}$$

are required to ensure that U is unitary, where I_1 and I_2 are unit matrices. Again, unitary matrices need not be square.

However, only square matrices can be Hermitian. For the equality

$$H = H^\dagger,$$

requires that the rows of H are labelled by the same scheme as the rows of $H^\dagger$. But the rows of $H^\dagger$ are the complex conjugates of the columns of H, so that the rows of H are labelled by the same scheme as the columns of H.

APPENDIX B

ANTILINEAR OPERATORS

$\hat{A}$ is an antilinear operator if, for any vectors $|X>$ and $|Y>$ in the domain of $\hat{A}$, and any scalers c_1 and c_2,

$$\hat{A}[c_1|X\rangle + c_2|Y\rangle] = c_1^*(\hat{A}|X\rangle) + c_2^*(\hat{A}|Y\rangle). \tag{B.1}$$

Obviously, then,

$$c\hat{A} = \hat{A}c^*. \tag{B.2}$$

If $\hat{A}$ and $\hat{B}$ are antilinear operators, the product $(\hat{A}\hat{B})$ is a linear operator. For,

$$\begin{aligned}(\hat{A}\hat{B})[c_1|X\rangle + c_2|Y\rangle] &= \hat{A}[c_1^*(\hat{B}|X\rangle) + c_2^*(\hat{B}|Y\rangle)] \\ &= c_1(\hat{A}\hat{B}|X\rangle) + c_2(\hat{A}\hat{B}|Y\rangle)\end{aligned} \tag{B.3}$$

which shows (see Eq. (2.41a)) that $\hat{A}\hat{B}$ is linear. In general, a product of p linear and q antilinear operators, is linear or antilinear according as q is even or odd.

Many of the operations and properties discussed in connection with linear operators (Section 2.2) are valid in the case of antilinear operators also. In particular, the inverse $\hat{A}^{-1}$ of $\hat{A}$ is defined by

$$\hat{A}\hat{A}^{-1} = \hat{1} = \hat{A}^{-1}\hat{A}. \tag{B.4}$$

Since $\hat{1}$ is linear, it follows from (B.3) that $\hat{A}^{-1}$ is antilinear. As in the case of linear operators, the necessary and sufficient condition that the antilinear operator $\hat{A}$ has an inverse, is that $\hat{A}\Phi$ be unique for each vector Φ.

In the following, we will confine ourselves to a discussion of the important differences of antilinear operators from linear operators.

Eq. (B.2) represents one of these important differences. That is, whereas a scalar commutes with every linear operator, only a *real* scalar commutes with an antilinear operator. Thus, as far as antilinear operators are concerned, complex numbers could be regarded as operators. The other important difference concerns the scalar product involving the operator. In the case of a linear operator, we have the relation (see Eqs. (2.9a) and (2.54)),

$$(\mathbf{Y}, \hat{A}\mathbf{X})^* = (A\mathbf{X}, \mathbf{Y}) = (\mathbf{X}, \hat{A}^\dagger\mathbf{Y}). \tag{B.5}$$

But, for antilinear $\hat{A}$, the corresponding relationship is

$$(\mathbf{Y}, \hat{A}\mathbf{X})^* = (\hat{A}\mathbf{Y}, \mathbf{X}) = (\hat{A}^\dagger\mathbf{X}, \mathbf{Y}) = (\mathbf{X}, \hat{A}^\dagger\mathbf{Y})^* \tag{B.6a}$$

or, in bra-ket notation;

$$[\langle Y \mid \{\hat{A} \mid X\rangle\}]^* = \{\langle Y \mid \hat{A}\} \mid X\rangle =$$

$$= \{\langle X \mid \hat{A}^\dagger\} \mid Y\rangle = [\langle X \mid \{\hat{A}^\dagger \mid Y\rangle\}]^* \tag{B.6b}$$

We see that $\hat{A}$ operating on one vector in a scalar product is equivalent to $\hat{A}^\dagger$ operating on the other vector (cf. Eq. (2.104)). Eq. (B.6a) also defines the Hermitian conjugate $\hat{A}^\dagger$ of $\hat{A}$. From this definition, we have, for three antilinear operators $\hat{A}, \hat{B}, \hat{C}$, the result,

$$(\hat{A}\hat{B}\hat{C}\mathbf{Y}, \mathbf{X}) = ((\hat{A}\hat{B}\hat{C})^\dagger \mathbf{X}, \mathbf{Y}), \tag{B.7a}$$

since $(\hat{A}\hat{B}\hat{C})$ is an antilinear operator. But, since $(\hat{A}\hat{B})$ and $(\hat{B}\hat{C})$ are linear operators, we get using (B.5) and (B.6a),

$$(\hat{A}\hat{B}\hat{C}\mathbf{Y}, \mathbf{X}) = (\hat{C}\mathbf{Y}, (\hat{A}\hat{B})^\dagger \mathbf{X})$$

$$= (\hat{C}^\dagger (\hat{A}\hat{B})^\dagger \mathbf{X}, \mathbf{Y}), \tag{B.7b}$$

and

$$(\hat{A}\hat{B}\hat{C}\mathbf{Y}, \mathbf{X}) = (\hat{A}^\dagger \mathbf{X}, \hat{B}\hat{C}\mathbf{Y})$$

$$= ((\hat{B}\hat{C})^\dagger \hat{A}^\dagger \mathbf{X}, \mathbf{Y}). \tag{B.7c}$$

Equality of all the three expressions (B.7a-c) requires that

$$(\hat{A}\hat{B}\hat{C})^\dagger = \hat{C}^\dagger \hat{B}^\dagger \hat{A}^\dagger, \tag{B.8}$$

which relationship is the same as for linear operators (Problem 2.7).

Note that, as a result of (B.6b), it is important to explicitly specify in a scalar product, whether an antilinear operator is operating on the ket vector to the right or on the bra vector to the left. Thus, $\langle X \mid \hat{A}\hat{B} \mid Y\rangle$ is not specific enough, but one can write,

$$\langle X \mid (\hat{A}\hat{B}) \mid Y\rangle = \{\langle X \mid \hat{A}\hat{B}\} \mid Y\rangle$$

$$= [\{\langle X \mid \hat{A}\} \{\hat{B} \mid Y\rangle\}]^*$$

$$= \langle X \mid \{\hat{A}\hat{B} \mid Y\rangle\}. \tag{B.9}$$

Antiunitary Operators

An antilinear operator $\hat{K}$ is said to be antiunitary if,

$$\hat{K}^\dagger = \hat{K}^{-1}$$

That is, if

$$\hat{K}\hat{K}^\dagger = \hat{K}^\dagger \hat{K} = \hat{1} \tag{B.10}$$

The product of an even number of antiunitary operators is unitary whereas the product of an odd number of antiunitary operators is anti-unitary.

A similarity transformation by an antiunitary operator also preserves the Hermitian character of a linear operator. This follows from Eqs. (2.57) and (B.8) which hold good both for linear and antilinear operators. However, from Eqs. (B.6a) and (B.10), we have,

$$(\mathbf{X},\mathbf{Y}) \to (\overline{\mathbf{X}},\overline{\mathbf{Y}}) = (\overline{\mathbf{X}},\hat{K}\mathbf{Y}) = (\mathbf{Y},\hat{K}^{\dagger}\overline{\mathbf{X}})$$

$$= (Y,\hat{K}^{\dagger}\hat{K}\mathbf{X}) = (\mathbf{Y},\mathbf{X}) = (\mathbf{X},\mathbf{Y})^{*}, \tag{B.11}$$

when $\hat{K}$ is antiunitary. Thus, only the absolute value of the scalar product is preserved under an antiunitary transformation:

$$|(\overline{\mathbf{X}},\overline{\mathbf{Y}})| = |(\mathbf{X},\mathbf{Y})^{*}| = |(\mathbf{X},\mathbf{Y})| \tag{B.11a}$$

In particular, the norm of a vector is invariant under antiunitary transformations.

Thus, the absolute value of the scalar product of vectors is preserved both by unitary (Eq. (2.62)) and antiunitary transformations. Conversely, *a transformation which preserves the absolute value of the scalar product is either unitary or antiunitary*[1].

The properties of an antiunitary transformation represented by the antiunitary operator $\hat{K}$, could be summarised as follows:

(i) A ket-vector $|X\rangle$ is transformed into

$$|\overline{X}\rangle = \hat{K}|X\rangle \tag{B.12a}$$

This follows from the definition (B.1) of antilinear operators.

(ii) A bra-vector $\langle X|$ is transformed into

$$\langle\overline{X}| = \langle X|\hat{K}^{\dagger} \tag{B.12b}$$

This can be proved using Eqs. (B.6a) and (B.11). From (B.11), we have,

$$\langle\overline{X}|\overline{Y}\rangle = \langle\overline{X}|\{\hat{K}|Y\rangle\} = \langle X|Y\rangle^{*}$$

But by Eq. (B.6a), we have,

$$\langle\overline{X}|\{\hat{K}|Y\rangle\} = [\{\langle\overline{X}|\hat{K}\}|Y\rangle]^{*}.$$

Therefore, $\langle\overline{X}|\hat{K} = \langle X|$,

which requires Eq. (B.12b)[2].

(iii) A linear operator $\hat{B}$ transforms into

$$\overline{\hat{B}} = \hat{K}\hat{B}\hat{K}^{\dagger}. \tag{B.13a}$$

This could be deduced, as follows:

Let $$|Y\rangle = \hat{B}|X\rangle.$$

Applying the antiunitary transformation $\hat{K}$, we get,

$$\hat{K}|Y\rangle = \hat{K}\hat{B}|X\rangle$$

$$= (\hat{K}\hat{B}\hat{K}^{\dagger})\hat{K}|X\rangle.$$

That is,

$$|\overline{Y}\rangle = \overline{\hat{B}}|\overline{X}\rangle.$$

---

1. For a proof, see A. Messiah, *Quantum Mechanics* (North Holland, Amsterdam 1961), Chapter XV, Section 2.
2. Note that a similar procedure using Eq. (B.5) yields, in the case of a unitary operator $\hat{U}$, the result,

$$\langle\overline{X}| = \langle X|\hat{U},$$

in agreement with Eq. (2.105).

Hence the result.

(iv) A complex number c transforms into

$$\overline{c} = \hat{K} c \hat{K}^{\dagger} = c^{*}, \tag{B.13b}$$

where Eqs. (B.2) and (B.10) have been used.

The proof is identical with that of Eq. (B.13a), with c replacing $\hat{B}$. The result (B.13b) confirms our earlier statement, in connection with Eq. (B.2), that complex numbers behave like linear operators under anti-linear operations. We see that Eq. (B.11) is actually only a special case of (B.13b), since the scalar product is a complex number. Similarly, it follows that the matrix representing $\hat{B}$ would be transformed into its complex conjugate under an antiunitary transformation. That is,

$$B_{jk} \equiv \langle u_j \mid \hat{B} \mid u_k \rangle \rightarrow B_{jk}^{*}. \tag{B.13c}$$

For,

$$\begin{aligned}
\langle u_j \mid \hat{B} \mid u_k \rangle &\rightarrow \langle \overline{u}_j \mid \overline{\hat{B}} \mid \overline{u}_k \rangle \\
&= \{\langle u_j \mid \hat{K}^{\dagger}\} (\hat{K} \hat{B} \hat{K}^{\dagger}) \{\hat{K} \mid u_k\rangle\} \\
&= \{\langle u_j \mid \hat{K}^{\dagger}\} \{\hat{K} \hat{B} \hat{K}^{\dagger} \hat{K} \mid u_k\rangle\} \\
&= [\{\langle u_j \mid \hat{K}^{\dagger} \hat{K}\}] \{\hat{B} \hat{K}^{\dagger} \hat{K} \mid u_k\rangle\}]^{*} \\
&= \langle u_j \mid \hat{B} \mid u_k \rangle^{*},
\end{aligned}$$

where use has been made of Eqs. (B.12a, b), (B.13a), (B.9), (B.6a) and (B.10).

It follows from (B.13a, b) that operator equations involving complex coefficients will be transformed, under antiunitary transformations, into the same equations with the coefficients replaced by their complex conjugates. Thus, the commutation relations,

$$[\hat{q}, \hat{p}] = i\hbar\, \hat{1} \quad \text{and} \quad [\hat{J}_x, \hat{J}_y] = i\hbar \hat{J}_z,$$

are transformed into

$$[\overline{\hat{q}}, \overline{\hat{p}}] = -i\hbar\, \hat{1}, \quad \text{and} \quad \left[\overline{\hat{J}}_x, \overline{\hat{J}}_y\right] = -i\hbar \overline{\hat{J}}_z. \tag{B.14}$$

Examples of Antiunitary Operators

The *complex conjugation operator* $\hat{K}_c$, which transforms a *c-number*[3] to its complex conjugate, is, obviously, an antiunitary operator since, by definition, it satisfies the equations (B.1), (B.13b,c) and (B.11). $\hat{K}_c$ also satisfies the relationship.

---

3. The term, "c-numbers" refers to ordinary numbers (real or complex) which obey a commutative algebra. This term was introduced by P.A.M. Dirac [*Proceedings of the Royal Society of London*(A) **110**, 561 (1926)] to distinguish such numbers from the operators of quantum mechanics (the "q-numbers") which obey a non-commutative algebra.

$$\hat{K}_c^2 = \hat{1}, \tag{B.15}$$

so that, from $\hat{K}_c^\dagger \hat{K}_c = \hat{1}$ (which follows from (B.11)), we get

$$\hat{K}_c^\dagger = \hat{K}_c. \tag{B.16}$$

The effect of $\hat{K}_c$ on a vector or on an operator, depends on the representation. Thus, in the representation defined by the basis $\{|u_k\rangle\}$,

$$\hat{K}_c |u_k\rangle = |u_k\rangle. \tag{B.17}$$

If $|X\rangle$ is an arbitrary vector, then,

$$|X\rangle = \sum_k |u_k\rangle \langle u_k | X\rangle$$

and

$$\hat{K}_c |X\rangle = \sum_k |u_k\rangle \langle u_k | X\rangle^*. \tag{B.18a}$$

That is, $\hat{K}_c |X\rangle$ is a vector whose components are the complex conjugates of the components of $|X\rangle$. On the other hand, in a representation in which $|X\rangle$ itself is a basis vector, we have[4],

$$\hat{K}_c |X\rangle = |X\rangle. \tag{B.18b}$$

Similarly, suppose $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$ are, respectively, the operators corresponding to the position vector and the momentum of a particle. Then, in the co-ordinate representation, $\hat{\mathbf{r}}$ is real whereas $\hat{\mathbf{p}}$ is pure imaginary [Eq. (3.18)], so that, we have,

$$\left.\begin{aligned} \hat{K}_c \hat{\mathbf{r}} \hat{K}_c^\dagger &= +\hat{\mathbf{r}}, \\ \hat{K}_c \hat{\mathbf{p}} \hat{K}_c^\dagger &= -\hat{\mathbf{p}}. \end{aligned}\right\} \text{ co-ordinate representation.} \tag{B.19a}$$

But, in the momentum representation, $\hat{\mathbf{p}}$ is real and $\hat{\mathbf{r}}$ is pure imaginary [Eq. (3.18')], so that,

$$\left.\begin{aligned} \hat{K}_c \hat{\mathbf{r}} \hat{K}_c^\dagger &= -\hat{\mathbf{r}}, \\ \hat{K}_c \hat{\mathbf{p}} \hat{K}_c^\dagger &= \hat{\mathbf{p}}. \end{aligned}\right\} \text{ momentum representation.} \tag{B.19b}$$

In either representation,

$$\hat{K}_c \hat{\mathbf{J}} \hat{K}_c^\dagger = -\hat{\mathbf{J}}, \tag{B.19c}$$

where $\hat{\mathbf{J}}$ is the angular momentum operator.

An arbitrary antilinear operator $\hat{A}$ could be written as the product of $\hat{K}_c$ and a linear operator. For,

$$\hat{A} = (\hat{A}\hat{K}_c)\hat{K}_c = \hat{A}_l \hat{K}_c, \tag{B.20a}$$

where $\hat{A}_l = (\hat{A}\hat{K}_c)$, is a linear operator (being the product of two antilinear operators). Also,

---

4. Eqs. (B.17) and (B.18b) follow from the fact that a basis vector is represented by a real (column) matrix [Eq. (2.112)] in a representation defined by a basis of which it is a member.

$$\hat{A} = \hat{K}_c(\hat{K}_c\hat{A}) = \hat{K}_c\hat{B}_l, \tag{B.20b}$$

with,
$$\hat{B}_l = \hat{K}_c\hat{A} = \hat{K}_c(\hat{A}\hat{K}_c)\hat{K}_c^\dagger = \overline{\hat{A}}_l. \tag{B.21}$$

Another example of an antilinear operator, is the *time-reversal operator* $\hat{\mathcal{T}}$, defined by,

$$\hat{\mathcal{T}}\hat{\mathbf{r}}\hat{\mathcal{T}}^\dagger = \hat{\mathbf{r}};\ \hat{\mathcal{T}}\hat{\mathbf{p}}\hat{\mathcal{T}}^\dagger = -\hat{\mathbf{p}}. \tag{B.22}$$

The transformation represented by $\hat{\mathcal{T}}$, obviously, satisfies Eq. (B.14) and is, hence, antiunitary. A fuller discussion on $\hat{\mathcal{T}}$ will be found in Section 6.2E.

APPENDIX C

FOURIER SERIES AND FOURIER TRANSFORMS

C.1 FOURIER SERIES

If $\psi(x)$ is a function which is single-valued, finite, has finite number of discontinuities and finite number of maxima and minima[1] in the interval $-\frac{L}{2} \le x \le \frac{L}{2}$, then according to *Fourier's Theorem*, it can be expanded in a *Fourier series*:

$$\psi(x) = \sum_{n=-\infty}^{+\infty} a_n e^{inkx}, \tag{C.1}$$

where,

$$a_n = \frac{1}{L}\int_{-L/2}^{L/2} \psi(x) e^{-inkx} dx, \tag{C.2}$$

and

$$k = \frac{2\pi}{L} \tag{C.3}$$

Eq. (C.2) follows from the *orthonormality* of the functions $\frac{1}{\sqrt{L}} e^{inkx}$. Writing, $\phi_n(x) = \frac{1}{\sqrt{L}} e^{inkx}$, we have,

$$(\phi_n, \phi_m) \equiv \int_{-L/2}^{L/2} \phi_n^*(x)\phi_m(x)dx = \frac{1}{L}\int_{-L/2}^{L/2} \exp[i(m-n)kx]dx = \delta_{m,n} \tag{C.4}$$

In terms of the ϕ_n, Eqs. (C.1) and (C.2) reads (cf. Eqs. (2.30) and (2.31)):

$$\psi(x) = \sum_{n=-\infty}^{+\infty} a'_n \phi_n(x) \tag{C.1a}$$

---

1. These conditions are called *Dirichlet conditions*, and a function which satisfies these conditions may be called *piece-wise regular*. Whereas the Dirichlet conditions are sufficient to make Eq. (C.1) valid, they are not all necessary.

with $$a'_n = \sqrt{L}\, a_n = (\phi_n, \psi). \tag{C.2a}$$

Thus, the Fourier series expansion could be regarded as the expansion of the 'vector' $\psi(x)$ in an infinite-dimensional Hilbert space, in terms of the basis vectors ϕ_n [see Eqs. (2.30) and (2.33)].

L is called the *period*. Fourier's theorem can be, however, applied even in the case of a function $F(x)$ which is not periodic, but which is known only within the interval $-L/2 \le x \le L/2$ and which satisfies the above *Dirichlet conditions* within this interval. In this case, $\psi(x)$ given by (C.1) should be understood as a periodic function which coincides with $F(x)$ in the interval $-L/2 \le x \le L/2$ (see Fig. C.1), but could differ from $F(x)$ outside this interval.

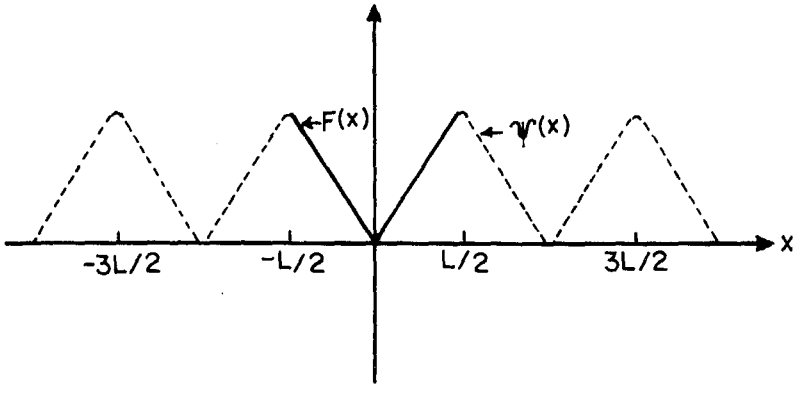


Fig. C.1. A function $F(x)$ [solid curve] and its Fourier series representation $\psi(x)$ [dotted curve].

Eq. (C.1) is the *complex* Fourier series. The *real* Fourier series can be obtained by writing $e^{inkx} = \cos(nkx) + i \sin(nkx)$. We get,

$$\psi(x) = \frac{b_0}{2} + \sum_{n=1}^{\infty} (b_n \cos nkx + c_n \sin nkx), \tag{C.5}$$

with $$b_n = a_n + a_{-n} = \frac{2}{L}\int_{-L/2}^{L/2} \psi(x) \cos nkx\, dx, \tag{C.6a}$$

$$c_n = i(a_n - a_{-n}) = \frac{2}{L}\int_{-L/2}^{L/2} \psi(x) \sin nkx\, dx. \tag{C.6b}$$

Changing the variable x to $t = kx = \left(\frac{2\pi}{L}\right)x$ in Eq. (C.5), we get an alternate expression for the *real* Fourier series:

$$f(t) = \frac{b_0}{2} + \sum_{n=1}^{\infty} [b_n \cos nt + c_n \sin nt], \tag{C.5'}$$

with[2]

2. The interval $-\pi$ to $+\pi$ in Eqs. (C.6a') and (C.6b') could be shifted to 0 to 2π. Correspondingly in Eq. (C.2), the interval would be from 0 to L.

$$b_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\cos nt\,dt, \tag{C.6a'}$$

$$c_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\sin nt\,dt, \tag{C.6b'}$$

where, $f(t) = \psi(x)$.

Parseval's Formula

Consider the function,

$$\chi_N(x) = \sum_{n=-N}^{+N} d_n \exp(inkx) \tag{C.7}$$

The choice of the coefficients d_n such that the quantity,

$$\epsilon_N = \int_{-L/2}^{L/2} |\psi(x) - \chi_N(x)|^2\,dx, \tag{C.8}$$

is minimum, makes χ_N an *approximation-in-the-mean* to $\psi(x)$.

Now,

$$\int_{-L/2}^{L/2} |\psi(x) - \chi_N(x)|^2\,dx = \int_{-L/2}^{L/2} |\psi(x)|^2\,dx + L\sum_{n=-N}^{+N}\{|d_n|^2 - a_n^* d_n - d_n^* a_n\}$$

$$= \int_{-L/2}^{L/2} |\psi(x)|^2\,dx - L\sum_{n=-N}^{+N} |a_n|^2 + L\sum_{n=-N}^{+N} |d_n - a_n|^2. \tag{C.9}$$

Thus, ϵ_N is minimum when $d_n = a_n$. In other words, the Fourier coefficients make every partial sum in (C.1) an approximation in-the-mean to $\psi(x)$.

From (C.9), we have, with the choice $d_n = a_n$,

$$\int_{-L/2}^{+L/2} |\psi(x)|^2\,dx - L\sum_{n=-N}^{+N} |a_n|^2 = \epsilon_N \tag{C.10}$$

Taking the limit, $N \to \infty$ in (C.10), we get, since $\lim_{N\to\infty} \chi_N = \psi$, so that $\lim_{N\to\infty} \epsilon_N = 0$,

$$\frac{1}{L}\int_{-L/2}^{L/2} |\psi(x)|^2\,dx = \sum_{n=-\infty}^{+\infty} |a_n|^2. \tag{C.11}$$

This is known as *Parseval's formula*. This formula is an expression of the convergence property of the partial sums χ_N and hence of the Fourier series. Alternatively (see Section 2.1, *completeness*), it expresses the completeness of the linear vector space spanned by the basis vectors ϕ_n, (C.11) being the norm of the vector ψ (cf. Eq. 2.25)).

The Fourier series enables us to represent a function with discontinuities (hence a function which is not analytic) by a function which is analytic. As an example, consider the function (cf. (C.5')),

$$f(x) = \frac{h}{2}, \text{ for } 0 < x < \pi,$$

$$= -\frac{h}{2}, \text{ for } -\pi < x < 0, \tag{C.12}$$

which represents a square wave (Fig. C.2).

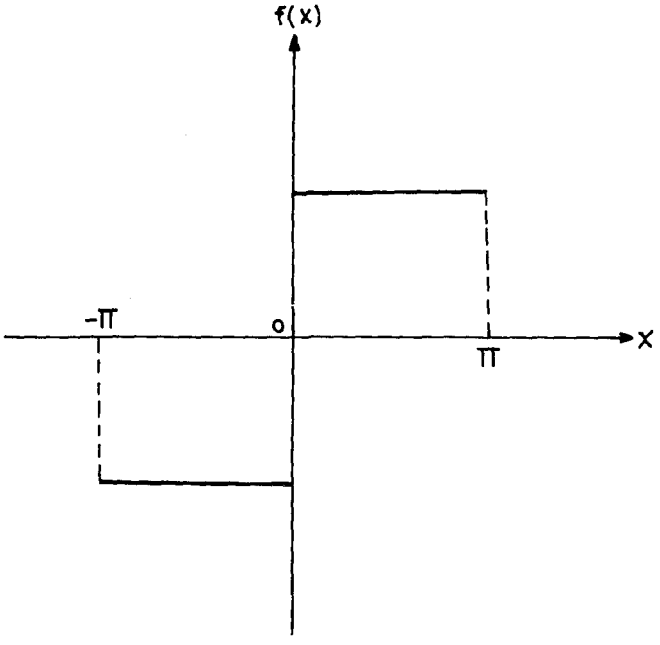


Fig. C.2. The square wave of Eq. (C.12).

Using (C.6a′) and (C.6b′), we get,

$$b_n = 0, c_n = \frac{2h}{\pi n},$$

so that,

$$f(x) = \frac{2h}{\pi} \sum_{n=1}^{\infty} \frac{\sin nx}{n}, \; n \text{ odd}$$

$$= \frac{2h}{\pi} \sum_{l=0}^{\infty} \frac{\sin(2l+1)x}{2l+1} \tag{C.13}$$

In Fig. (C.3), we show the partial sums,

$$X_N(x) = \frac{2h}{\pi} \sum_{l=0}^{N-1} \frac{\sin(2l+1)x}{2l+1}, \tag{C.14}$$

for $N = 20, 40, 60, 80$ and 100, near the discontinuity at $x = 0$. We see that whereas X_N progressively approaches $f(x)$ as N increases, it consistently overshoots $f(x)$ in the vicinity of the discontinuity at $x = 0$. This is known as *Gibb's phenomenon*[3].

3. See, Arfken, G. *Mathematical Methods for Physicists* (Academic Press, New York 1970), II Edition, section 14.5.

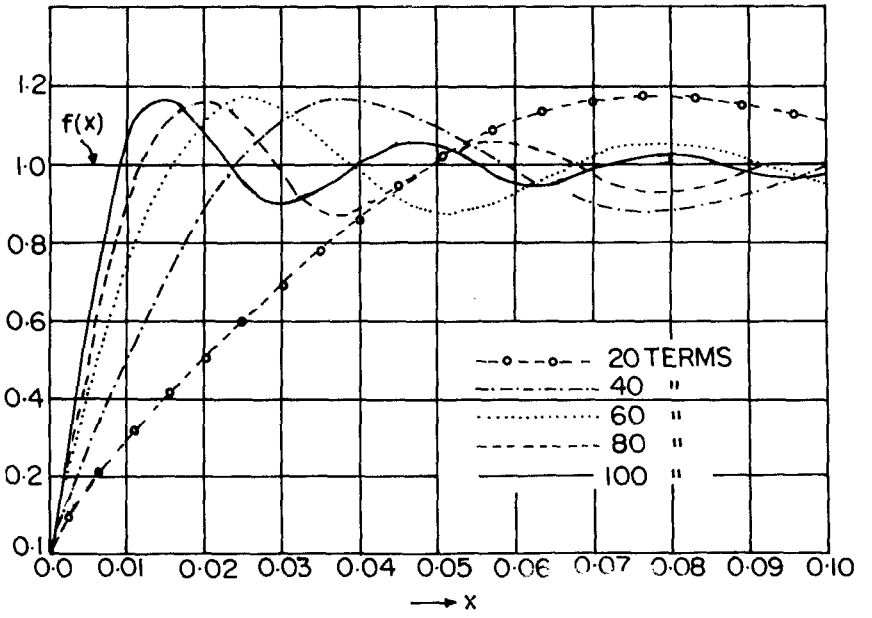


Fig. C.3. Gibb's phenomenon.

C.2 FOURIER TRANSFORMS

When the function $\psi(x)$ is not periodic (that is, when the period $L = \infty$), the Fourier series representation (C.1) of $\psi(x)$ has to be suitably modified as the expressions (C.2) and (C.3) become meaningless when $L = \infty$. The required modification could be done as follows:

Substituting from (C.2) in (C.1), we have,

$$\psi(x) = \sum_{n=-\infty}^{+\infty} \frac{1}{L} \int_{-L/2}^{L/2} \psi(x') \exp\,[ink(x-x')]\,dx'$$

where, we have changed the variable of integration to x' from x in (C.2). Again, using relationship (C.3) for k, we get,

$$\psi(x) = \sum_{n=-\infty}^{+\infty} \frac{1}{L} \int_{-L/2}^{L/2} \psi(x') \exp\left[\frac{2\pi i n}{L}(x-x')\right] dx' \qquad \text{(C.15)}$$

Letting $L \to \infty$ and writing $\frac{1}{L} = \Delta s$, Eq. (C.15) reduces to

$$\psi(x) = \sum_{n=-\infty}^{+\infty} \Delta s \int_{-L/2}^{+L/2} \psi(x') \exp\,(i2\pi n \Delta s(x-x'))dx' \qquad \text{(C.15a)}$$

Now,

$$\sum_{n=-\infty}^{+\infty} f(n\Delta s)\Delta s\,\bar{\Delta} s \to 0 \int_{-\infty}^{+\infty} f(s)ds, \qquad \text{(C.16)}$$

so that,

$$\sum_{n=-\infty}^{+\infty} \Delta s \cdot \exp(i2\pi n \Delta s(x-x'))\tilde{\Delta} s \to 0 \int_{-\infty}^{+\infty} \exp(i2\pi s(x-x'))ds \tag{C.17}$$

Substituting (C.17) in (C.15a), we get,

$$\psi(x) = \int_{-\infty}^{+\infty} \{\int_{-\infty}^{+\infty} \psi(x') \exp(-i2\pi s x')dx'\} \exp(i2\pi s x)ds \tag{C.18}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \phi(k) \exp(+ikx)dk, \tag{C.18a}$$

where

$$\phi(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \psi(x) \exp(-ikx)dx, \tag{C.18b}$$

with

$$k = 2\pi s. \tag{C.3'}$$

Eq. (C.18) is the *Fourier Integral* representation[4] of $\psi(x)$. In addition to the Dirichlet conditions, the existence of the integral $\int_{-\infty}^{+\infty} \psi(x)\,dx$ is also required for the validity of this expression.

$\phi(k)$ given by Eq. (C.18b) is called the *Fourier Transform* of $\psi(x)$ and could be symbolically written as,

$$\phi(k) = \mathcal{F}\{\psi(x)\} \tag{C.18b'}$$

$\psi(x)$ could be, then, regarded as the Inverse Fourier transform of $\phi(k)$:

$$\psi(x) = \mathcal{F}^{-1}\{\phi(k)\} \tag{C.18a'}$$

Comparing (C.18a) with (C.1), we see that $\phi(k)$ is the *amplitude* of the harmonic component of 'wave number' k in the resolution of $\psi(x)$ into harmonic waves.

The real and imaginary parts of Eq. (C.18b):

$$\phi_c(k) = \sqrt{\frac{1}{2\pi}} \int_0^{\infty} \psi_e(x) \cos kx dx \tag{C.18c}$$

$$\phi_s(k) = \sqrt{\frac{1}{2\pi}} \int_0^{\infty} \psi_0(x) \sin kx dx, \tag{C.18d}$$

are, respectively, known as the *Fourier cosine* and the *Fourier sine* transforms. Here,

$$\psi_e(x) = \psi(x) + \psi(-x) \tag{C.19a}$$

---

4. If k is replaced by αk, then the constant $\frac{1}{\sqrt{2\pi}}$ should be replaced by $\frac{1}{\sqrt{2\pi\alpha}}$.

and

$$\psi_0(x) = \psi(x) - \psi(-x) \tag{C.19b}$$

As an example, consider the pulse,

$$\psi(x) = A, \; x < |a|$$

$$= 0, \; x > |a| \tag{C.20}$$

From (C.18b), we get,

$$\phi(k) = \frac{1}{\sqrt{2\pi}} A \int_{-a}^{+a} \exp(-ikx) dx$$

$$= \frac{A}{\sqrt{2\pi}} (e^{-ika} - e^{+ika})/(-ik)$$

$$= \sqrt{\frac{2}{\pi}} Aa \left(\frac{\sin ka}{ka} \right), \tag{C.21}$$

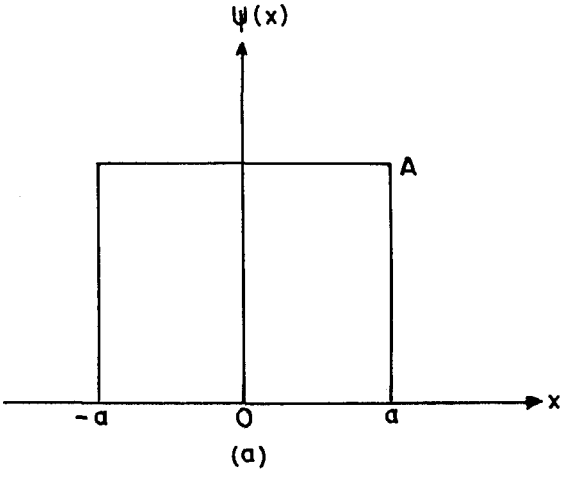


Fig. C.4.(a) The function $\psi(x)$.

which could be obtained also from (C.18c), noting that $\psi(x)$ is an even function of x. Both $\psi(x)$ and $\phi(k)$ are plotted in Fig. (C.4). We note that $\phi(k)$ is appreciable only within an interval of k given by

$$\Delta k \approx \frac{\pi}{a} \tag{C.22a}$$

whereas $\psi(x)$ is non-zero within an interval of x given by

$$\Delta x = 2a$$

Thus, the product,

$$\Delta x \cdot \Delta k \approx 2\pi \tag{C.23}$$

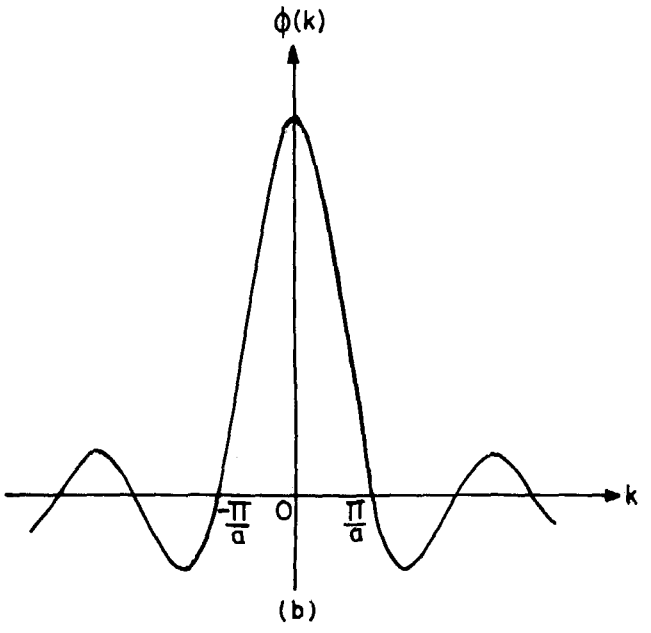


Fig. C.4.(b) The Fourier transform of $\psi(x)$.

This is a relationship generally valid for a function and its Fourier transform. Thus, if $\psi(x)$ is a Gaussian function:

$$\psi(x) = \left(\frac{1}{a\sqrt{\pi}}\right)^{1/2} \exp\left(-\frac{1}{2}(x^2/a^2)\right) \tag{C.24a}$$

then, $\phi(k)$ is also a Gaussian function,

$$\phi(k) = \left(\frac{a}{\sqrt{\pi}}\right)^{1/2} \exp\left(-\frac{1}{2}a^2k^2\right), \tag{C.24b}$$

so that,

$$\Delta x \sim a, \Delta_k \sim \frac{1}{a}$$

and

$$\Delta x \cdot \Delta k \sim 1. \tag{C.23a}$$

The importance of this relationship in quantum mechanics, arises from the possibility of identifying it with the uncertainty relationship of Heisenberg (see, relationship (3.29a)). The expression,

$$\psi(x,t) = \frac{1}{\sqrt{2\pi}} \int_{\Delta k} \phi(k) \exp\left[i(kx - \omega t)\right] dk \tag{C.25a}$$

or, its 3-dimensional counterpart,

$$\psi(\mathbf{r},t)=\left(\frac{1}{\sqrt{2\pi}}\right)^3\int_{\Delta\mathbf{k}}\phi(\mathbf{k})\exp\left[i(\mathbf{k}\cdot\mathbf{r}-\omega t)\right]d^3\mathbf{k} \tag{C.25b}$$

where ω is the angular frequency, represents a travelling *wave packet* in configuration space, whereas the Fourier transform

$$\phi(k)=\frac{1}{\sqrt{2\pi}}\int_{\Delta x}\psi(x,t)\exp\left[-i(kx-\omega t)\right]dx \tag{C.26a}$$

or,

$$\phi(\mathbf{k})=\left(\frac{1}{\sqrt{2\pi}}\right)^3\int_{\Delta\mathbf{r}}\psi(\mathbf{r},t)\exp\left[-i(\mathbf{k}\cdot\mathbf{r}-\omega t)\right]d^3\mathbf{r} \tag{C.26b}$$

represents the same wave packet in the wave number space. From the point of view of particles, if $\psi(\mathbf{r},t)$ is the wave function of a system of particles in configuration space, then, $\phi(\mathbf{k})$ is the corresponding wave function in momentum space (taking into account the de Bröglie relationship $\mathbf{p}=\hbar\mathbf{k}$). Relationship (C.23a), then, is equivalent to

$$\Delta x\cdot\Delta p_x\sim\hbar,\ \text{cyclic} \tag{C.23b}$$

Δx being the 'spread' in the x-positions of the particles and Δp_x the spread in the x-components of the momenta of the particles.

The wave packet (C.25a) can also be expressed as

$$\psi(x,t)=\frac{1}{\sqrt{2\pi}}\int_{\Delta\omega}\chi(\omega)\exp\left(i(kx-\omega t)\right)d\omega, \tag{C.27a}$$

with

$$\chi(\omega)=\frac{1}{\sqrt{2\pi}}\int_{\Delta t}\psi(x,t)\exp\left(-i(kx-\omega t)\right)dt, \tag{C.27b}$$

such that

$$\Delta\omega\cdot\Delta t\sim 1, \tag{C.28a}$$

or, using the Planck-Einstein relationship $E=\hbar\omega$,

$$\Delta E\cdot\Delta t\sim\hbar. \tag{C.28b}$$

The meaning of (C.28a) is that if the wave packet is a super-position of harmonic waves with a spread in the angular frequencies equal to $\Delta\omega$, then the time Δt taken by the wave packet to pass a fixed point, say x_0 is of the order of $(1/\Delta\omega)$. The interpretation of (C.28b) as a time-energy uncertainty relationship analogous to (C.23b) is, however, beset with difficulties (see section 3.2: *The time energy uncertainty relationship*).

APPENDIX D

DIRAC DELTA FUNCTION

The Dirac delta function, $\delta(x)$, was introduced by P.A.M. Dirac in order to treat eigenvectors belonging to the continuous eigenvalues of linear operators (for example, the momentum eigenfunction $u_k(\mathbf{r}) = C\, e^{i\mathbf{k}.\mathbf{r}}$) on the same footing as those belonging to the discrete eigenvalues. In the case of the discrete spectrum, the orthonormality of the eigenvectors $\{u_k(x)\}$, is expressed by the relation (we are restricting ourselves to one dimension, for the sake of simplicity),

$$\int_{-\infty}^{+\infty} u_k^*(x) u_l(x) dx = \delta_{k,l}, \tag{D.1}$$

where $\delta_{k,l}$ is the *Kronecker* delta function, defined as

$$\delta_{k,l} = 1, \quad \text{for } k = l, \tag{D.2}$$

$$= 0, \quad \text{for } k \neq l.$$

When the spectrum is continuous (k is a continuous variable), the normalization of the eigenvectors are expressed by Eq. (D.1) with the Kronecker delta function replaced by the Dirac delta function:

$$\int_{-\infty}^{+\infty} u_k^*(x) u_l(x) dx = \delta(k - l), \tag{D.3}$$

where, in analogy with (D.2), $\delta(k-l)$ is defined as

$$\delta(k-l) = 0, \quad \text{for } k \neq l. \tag{D.4a}$$

However, being a continuous function of $(k-l)$, $\delta(k-l)$ cannot be defined as being equal to unity for $k = l$. The behaviour of $\delta(k-l)$ for $k = l$ could be inferred from the following consideration. In the case of a discrete spectrum an arbitrary wave function $\psi(x)$ is given, in terms of the u_k's, by

$$\psi(x) = \sum_k f_k u_k(x). \tag{D.5a}$$

Using Eq. (D.1), we obtain,

$$\int_{-\infty}^{+\infty} u_l^*(x) \psi(x) dx = \sum_k f_k \int_{-\infty}^{+\infty} u_l^*(x) u_k(x) dx$$

$$= \sum_k f_k \delta_{kl} = f_l \tag{D.5b}$$

The relationship corresponding to (D.5a), when the spectrum is continuous, i.

$$\psi(x) = \int_{-\infty}^{+\infty} f(k) u_k(x) dk, \tag{D.5b}$$

where, $f(k)$, is a continuous function of k. Multiplying both sides of (D.5b) by $u_l^*(x)$ and integrating over x, we have, using Eq. (D.3),

$$\int_{-\infty}^{+\infty} u_l^*(x)\psi(x)dx = \int_{-\infty}^{+\infty} f(k)\delta(k-l)dk. \tag{D.5c}$$

In order to agree with (D.3a), we should have,

$$\int_{-\infty}^{+\infty} f(k)\delta(k-l)dk = f(l). \tag{D.4b}$$

This equation, in addition to implying Eq. (D.4a), defines $\delta(k-l)$ for $k = l$ as well. Therefore, Eq. (D.4b) could be taken as the definition of the Dirac delta function. Substituting $f(k) = c$ (a constant), in (D.4b), we get,

$$\int_{-\infty}^{+\infty} \delta(k-l)dk = 1. \tag{D.4c}$$

Thus, $\delta(x)$ could be thought of as a function which is zero everywhere except in the neighbourhood of $x = 0$ where it is so large that the area enclosed by the curve $\delta(x)$ and the x-axis is unity. We see that, viewed as a function of x, the behaviour of $\delta(x)$ is rather 'peculiar'. It is, however, possible to provide a proper mathematical basis to the Dirac delta function within the framework of *distribution theory* where it turns out that $\delta(x)$ is not a function, but is a *functional*[1]. The definition (D.4b) is sufficient as far as the use of $\delta(x)$ in quantum mechanics is concerned. Therefore, we will regard $\delta(x)$ as any function of x that satisfies Eq. (D.4b) or, equivalently, Eqs. (D.4a) and (D.4c).

Representation of $\delta(x)$

Any function of x that satisfies either Eq. (D.4b) or Eqs. (D.4a) and (D.4c), provides a representation of $\delta(x)$. Thus, comparing the Fourier integral formula (Eq. (C.18)),

$$\psi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \psi(x')e^{-ik(x'-x)}dx'dk,$$

with (cf. Eq. (D.4b)),

$$\psi(x) = \int_{-\infty}^{+\infty} \psi(x')\delta(x'-x)dx',$$

we have,

$$\delta(x'-x) = \frac{1}{2\pi}\,\mathrm{Lt}_{L\to\infty}\int_{-L}^{L} e^{-ik(x'-x)}dk \tag{D.6a}$$

$$= \mathrm{Lt}_{L\to\infty}\frac{\sin L(x'-x)}{\pi(x'-x)}$$

1. See, Messiah, A. *Quantum Mechanics*, (North-Holland, Amsterdam 1961), Vol. I. Appendix A.

That is,

$$\delta(x) = \mathrm{Lt}_{L\to\infty} \frac{\sin Lx}{\pi x}. \tag{D.6b}$$

Some of the other useful representations are:

$$\delta(x) = \frac{1}{\pi}\,\mathrm{Lt}_{L\to\infty} \frac{1-\cos Lx}{Lx^2} \tag{D.6c}$$

$$= \frac{1}{\pi}\,\mathrm{Lt}_{\epsilon\to 0} \frac{\epsilon}{x^2+\epsilon^2} \tag{D.6d}$$

$$= \mathrm{Lt}_{\eta\to 0} \frac{\Theta(x+\eta)-\Theta(x)}{\eta} \tag{D.6e}$$

In (D.6e), $\Theta(x)$ is the *Heavyside step function*, defined by,

$$\Theta(x) = \begin{cases} 1, & \text{for } x > 0 \\ 0, & \text{for } x < 0. \end{cases} \tag{D.7}$$

Properties of $\delta(x)$

The following properties of $\delta(x)$ could be established using the defining equations (D.4a), (D.4b) or (D.4c):

$$\delta(x) = \delta(-x) \tag{D.8}$$

$$x\delta(x) = 0. \tag{D.9}$$

$$\delta(ax) = \frac{1}{|a|}\delta(x) \tag{D.10}$$

$$\delta(x^2-a^2) = \frac{1}{2a}[\delta(x-a)+\delta(x+a)] \tag{D.11}$$

$$f(x)\delta(x-a) = f(a)\delta(x-a) \tag{D.12}$$

$$\int \delta(x-y)\delta(y-a)dy = \delta(x-a) \tag{D.13}$$

$$\frac{d}{dx}\{\delta(x)\} = -\frac{d}{dx}\{\delta(-x)\}. \tag{D.14}$$

$$x\frac{d}{dx}\delta(x) = -\delta(x) \tag{D.15}$$

These equalities merely imply that both sides yield the same result when multiplied by a function $f(x)$ and integrated over x.

APPENDIX E

SPECIAL FUNCTIONS

In this Appendix, we will present the definition and properties of the polynomial solutions of certain second order, linear, homogeneous, differential equations of the type,

$$y'' + P(x)y' + Q(x)y = 0, \tag{E.1}$$

where the prime denotes differentiation with respect to x, and, P and Q are functions of x. According to *Fuch's theorem*, if x_0 is either an ordinary point, or a *non-essential singularity*[1] of the equation (E.1), then the equation has a solution in the form of an infinite series around x_0:

$$y = \sum_{\lambda=0}^{\infty} a_\lambda (x - x_0)^{k+\lambda}, \tag{E.2}$$

where k is a constant. a_λ and k can be determined by substituting (E.2) in Eq. (E.1) and then equating the coefficients of every power of x to zero. For details, the reader is referred to the book by G. Arfken[2]. We give below only a summary of the properties.

E.1 HERMITE POLYNOMIALS

These are solutions of the equation,

$$y'' - 2xy' + 2ny = 0, \tag{E.3}$$

(where n is a positive integer) and are given by

$$y_n(x) \equiv H_n(x) = \sum_{r=0}^{(s)} (-1)^r \frac{n!}{(n-2r)!r!} (2x)^{n-2r} \tag{E.4}$$

---

1. x_0 is an ordinary point of the equation (E.1) if $P(x_0)$ and $Q(x_0)$ are finite, whereas it is a singularity if $P(x_0)$ and $Q(x_0)$ are infinite. x_0 is a non-essential singularity if $(x-x_0)P(x_0)$ and $(x-x_0)^2 Q(x_0)$ are finite. Thus, for the equation

$$y'' - \frac{2x}{1-x^2}y' + \frac{\alpha}{1-x^2}y = 0,$$

the point $x = 0$ is an ordinary point, while $x = 1$ is a nonessential singularity.

2. Arfken, G. *Mathematical Methods for Physicists*, II Edition (Academic Press, New York, 1970) Chapters, 8, 11, 12 and 13.

$$= (-1)^n e^{x^2} \frac{d^n}{dx^n}\left(e^{-x^2}\right) \tag{E.4^2}$$

$$= \frac{n!}{2\pi i} \oint z^{-n-1} \exp\,(x^2 - (z-x)^2) dz. \tag{E.4^3}$$

In Eq. (E.4^3), the contour of integration is a circle with the centre as the origin and z is a complex number. In (E.4^1), $(s) = \frac{n}{2}$, for n even and $(s) = \frac{n-1}{2}$, for n odd.

Generating Function

$$\exp\,[x^2 - (z-x)^2] = \sum_{n=0}^{\infty} H_n(x)\left(\frac{z^n}{n!}\right). \tag{E.5}$$

Recurrence Relations

$$H'_n = 2nH_{n-1}. \tag{E.6a}$$

$$xH_n = \frac{1}{2}H_{n+1} + nH_{n-1}. \tag{E.6b}$$

$$x^2 H_n = \frac{1}{4}H_{n+2} + \left(n + \frac{1}{2}\right)H_n + n(n-1)H_{n-2}. \tag{E.6c}$$

Symmetry

$$H_n(-x) = (-1)^n H_n(x). \tag{E.7}$$

The first few Hermite polynomials are given in Table E.1.

Table E. 1. Hermite Polynomials

n	$H_n(x)$
0	1
1	$2x$
2	$4x^2 - 2$
3	$8x^3 - 12x$
4	$16x^4 - 48x^2 + 12$
5	$32x^5 - 160x^3 + 120x$

Hermite Orthonormal Functions

The Hermite orthonormal function, $\phi_n(x)$, is given by

$$\phi_n(x) = \frac{1}{[\sqrt{\pi}\,2^n n!]^{1/2}} e^{-x^2/2} H_n(x). \tag{E.8}$$

ϕ_n satisfies the differential equation,

$$\phi''_n + (1 + 2n - x^2)\phi_n = 0, \tag{E.9}$$

and the orthonormal relationship,

$$\int_{-\infty}^{+\infty} \phi_n^*(x)\phi_m(x)dx \equiv (\phi_n, \phi_m) = \delta_{n,m}. \tag{E.10}$$

This relationship could be proved using (E.8) and (E.5). From (E.6a-c) and the definition (E.8) we have the recurrence relations.

$$x\phi_n(x) = \sqrt{\frac{n}{2}}\phi_{n-1} + \sqrt{\frac{n+1}{2}}\phi_{n+1}, \tag{E.11a}$$

$$x^2\phi_n(x) = \frac{1}{2}[\sqrt{n(n-1)}\,\phi_{n-2} + (2n+1)\phi_n + \sqrt{(n+1)(n+2)}\,\phi_{n+2}], \tag{E.11b}$$

$$\phi'_n(x) = \sqrt{\frac{n}{2}}\phi_{n-1} - \sqrt{\frac{n+1}{2}}\phi_{n+1}, \tag{E.11c}$$

$$\phi''_n(x) = [x^2 - (1+2n)]\phi_n$$

$$= \frac{1}{2}\sqrt{n(n+1)}\,\phi_{n-2} - (2n+1)\phi_n + \sqrt{(n+1)(n+2)}\,\phi_{n+2}]. \tag{E.11d}$$

From Eqs. (E.10) and (E.11 a-d), it follows that,

$$(\phi_m, x\phi_n) = \frac{1}{\sqrt{2}}[\sqrt{n}\,\delta_{m,n-1} + \sqrt{n+1}\,\delta_{m,n+1}], \tag{E.12a}$$

$$(\phi_m, x^2\phi_n) = \frac{1}{2}[\sqrt{n(n-1)}\,\delta_{m,n-2} + (2n+1)\delta_{m,n} + \sqrt{(n+1)(n+2)}\,\delta_{m,n+2}], \tag{E.12b}$$

$$(\phi_m, \phi'_n) = \sqrt{\frac{1}{2}}[\sqrt{n}\,\delta_{m,n-1} - \sqrt{n+1}\,\delta_{m,n+1}], \tag{E.12c}$$

$$(\phi_m, \phi''_n) = \frac{1}{2}[\sqrt{n(n-1)}\,\delta_{m,n-2} - (2n+1)\delta_{m,n} + \sqrt{(n+1)(n+2)}\,\delta_{m,n+2}] \tag{E.12d}$$

In Fig. E.1, we have plotted $\phi_n(x)$ against x, for $n = 0$ to 5. The $\phi_n(x)$ are, actually, the normalized wave-functions of the linear harmonic oscillator (See, Section 4.2A).

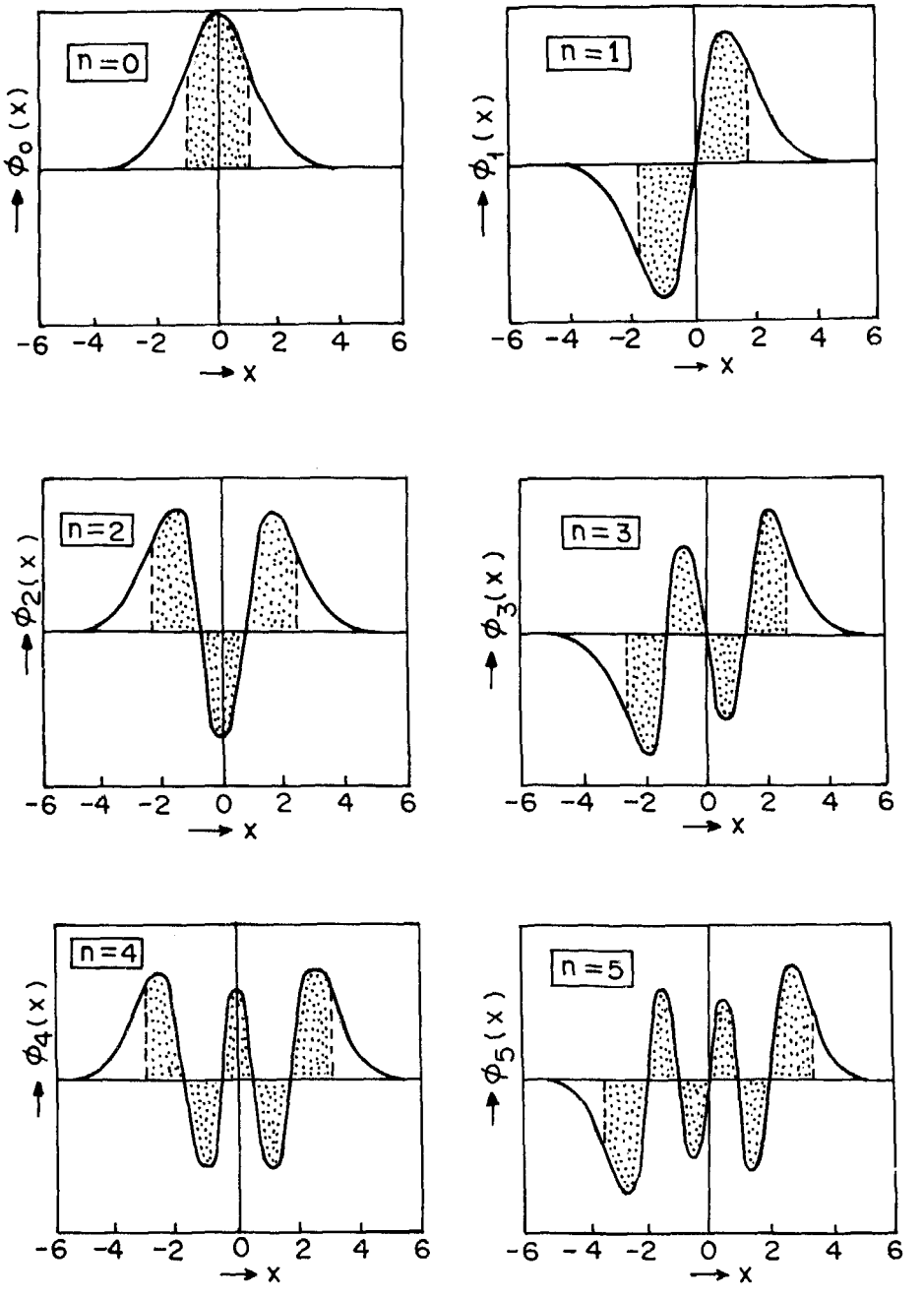


Fig. E.1. The Hermite orthonormal functions [Eq. (E.8)].

E.2 LAGUERRE POLYNOMIALS

The Laguerre polynomial $L_n(x)$ of degree n, is given by[3]

$$L_n(x) = \sum_{r=0}^{n} (-1)^r \frac{(n!)^2}{(r!)^2(n-r)!} x^r$$

$$= (-1)^n \left[x^n - \frac{n^2}{1!} x^{n-1} + \frac{n^2(n-1)^2}{2!} x^{n-2} \ldots \right] \quad \text{(E.13}^1\text{)}$$

$$= e^x \frac{d^n}{dx^n} (x^n e^{-x}) \quad \text{(E.13}^2\text{)}$$

$$= \frac{n!}{2\pi i} \oint \frac{e^{-xz/(1-z)}}{(1-z)z^{n+1}} dz \quad \text{(E.13}^3\text{)}$$

In Eq. (E.13^3), the contour of integration includes the origin, but excludes the point $z = 1$.

Generating Function

$$g(x,z) \equiv (1-z)^{-1} \{e^{-xz/(1-z)}\} = \sum_{n=0}^{\infty} L_n(x)(z^n/n!) \quad \text{(E.14)}$$

Recurrence Relations

One recursion relation is given by the differential equation satisfied by $L_n(x)$:

$$L''_n + \frac{1-x}{x} L'_n + \frac{n}{x} L_n = 0. \quad \text{(E.15a)}$$

The other can be derived from (E.14):

$$(1 + 2n - x)L_n - n^2 L_{n-1} - L_{n+1} = 0. \quad \text{(E.15b)}$$

The Laguerre polynomials for the lowest few values of n are given in Table E.2.

Table E.2 Laguerre Polynomials

n	$L_n(x)$
0	1
1	$-x + 1$
2	$x^2 - 4x + 2$
3	$-x^3 + 9x^2 - 18x + 6$
4	$x^4 - 16x^3 + 72x^2 - 96x + 24$
5	$-x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120$
	$L_n(0) = n!$

[3] Some definitions of $L_n(x)$ differ from the one adopted here by a factor of $1/n!$.

Associated Laguerre Polynomials

The Associated Laguerre polynomial $L_n^k(x)$ of degree $(n-k)$, is defined by[4]

$$L_n^k(x) = \frac{d^k}{dx^k}\{L_n(x)\},\ n \geq k, \tag{E.16}$$

and is a solution of the equation,

$$y'' + \frac{1+k-x}{x}y' + \frac{n-k}{x}y = 0, \tag{E.17}$$

where n and k are both positive integers. The *generating function* for L_n^k is given by

$$(-z)^k(1-z)^{-(k+1)}e^{-xz/(1-z)} = \sum_{n=k}^{\infty} L_n^k(x)\frac{z^n}{n!}, \tag{E.18}$$

from which the series expression for $L_n^k(x)$ follows:

$$L_n^k(x) = \sum_{r=0}^{n-k}(-1)^{r+k}\frac{(n!)^2}{(n-k-r)!\,(k+r)!\,r!}x^r. \tag{E.19}$$

From definitions (E.16) and (E.19), we have,

$$L_n^0(x) = L_n(x), \tag{E.19a}$$

and

$$L_n^n(x) = (-1)^n n!. \tag{E.19b}$$

Some of the associated Laguerre polynomials are listed in Table E.3.

Table E.3 Associated Laguerre Polynomials

n	k	$L_n^k(x)$
0	0	1
1	0	$-x+1$
	1	-1
2	0	x^2-4x+2
	1	$2x-4$
	2	2
3	0	$-x^3+9x^2-18x+6$
	1	$-3x^2+18x-18$
	2	$-6x+18$
	3	-6
4	3	$24x-96$
	4	24
5	4	$-120x+600$
	5	-120

4. The definition, $L_n^k(x) = \frac{d^k}{dx^k}\{L_{n+k}(x)\}$, is also used. In this case, L_n^k is a polynomial of degree n.

The Associated Laguerre Function

This is defined by

$$\mathcal{L}_{n,k}(x) = e^{-(x/2)} x^{(k-1)/2} L_n^k(x), \tag{E.20}$$

and satisfies the differential equation,

$$x\mathcal{L}''_{n,k} + 2\mathcal{L}'_{n,k} + \left\{ n - \frac{k-1}{2} - \frac{x}{4} - \frac{k^2-1}{4x} \right\} \mathcal{L}_{n,k} = 0. \tag{E.21}$$

Using the generating function (E.18) for $L_n^k(x)$ and the definition (E.20), it is possible to show:

$$\sum_{n,m=k}^{\infty} \frac{z_1^n z_2^m}{n!m!} I_{n,m}^{(p)}(k)$$

$$= [(1-z_1)(1-z_2)]^{p-1} \cdot \sum_{l=0}^{\infty} \frac{(k+p-1+l)!}{l!} (z_1 z_2)^{k+l}, \tag{E.22^1}$$

where

$$I_{n,m}^{(p)}(k) \equiv \int_0^\infty x^p \mathcal{L}_{n,k}(x) \mathcal{L}_{m,k}(x) dx, \; p \geq 0. \tag{E.23}$$

For $p \geq 1$, using the binomial expansion,

$$(1-z)^{p-1} = \sum_{r=0}^{p-1} (-1)^r \frac{(p-1)!}{(p-1-r)!r!} z^r,$$

Eq. (E.22^1) can be written as,

$$\sum_{n,m=k}^{\infty} \frac{z_1^n z_2^m}{n!m!} I_{n,m}^{(p)}(k)$$

$$= \sum_{r=0}^{p-1} \sum_{s=0}^{p-1} \sum_{l=0}^{\infty} (-1)^{r+s} \frac{\{(p-1)!\}^2 (k+p-1+l)!}{(p-1-r)!(p-1-s)!r!s!l!} \times z_1^{r+k+l} z_2^{s+k+l} \tag{E.22^2}$$

From this, we deduce,

$$I_{n,m}^{(p)}(k) = \sum_{r=0}^{p-1} \left\{ \frac{n!(p-1)!}{(p-1-r)!r!} \right\}^2 \cdot \frac{(p-1-r+n)!}{(n-k-r)!}; \; p \geq 1. \tag{E.24}$$

Explicit expressions for some of the $I_{n,m}^{(p)}$ are given below:

$$I_{n,m}^{(1)}(k) = \frac{(n!)^3}{(n-k)!} \delta_{n,m}, \tag{E.25a}$$

$$I_{n,m}^{(2)}(k) = \frac{(n!)^3}{(n-k)!} (2n-k+1), \text{ for } m = n \tag{E.25b}$$

$$= \frac{(n!)^3}{(n-k)!} \left[\frac{(n-k)}{n} \delta_{m,n-1} + (n+1)^2 \delta_{m,n+1} \right], \; m \neq n \tag{E.25c}$$

$$I^{(3)}_{n,m}(k) = \frac{(n!)^3}{(n-k)!}\,[6n(n-k+1)+k^2-3k+2], m=n, \tag{E.25d}$$

$$= \frac{(n!)^3}{(n-k)!}\left[\frac{2(n-k)(2n-k)}{n}\,\delta_{m,n-1}\right.$$

$$+ 2(n+1)^2(2n-k+2)\,\delta_{m,n+1}$$

$$\left.\frac{(n-k)(n-k-1)}{n(n-1)}\,\delta_{m,n-2} - (n+1)^2(n+2)^2\,\delta_{m,n+2}\right],\ m \neq n. \tag{E.25e}$$

E.3 LEGENDRE POLYNOMIALS

The Legendre polynomial $P_l(x)$ of degree l ($l = 0, 1, 2, \ldots + \infty$), is defined, for $-1 \le x \le 1$, by

$$P_l(x) = \sum_{k=0}^{s} (-1)^k \frac{(2l-2k)!}{2^n k!(l-k)!(l-2k)!} x^{l-2k}, \tag{E.26a}$$

$$= \frac{1}{2^l l!}\frac{d^l}{dx^l}(x^2-1)^l, \tag{E.26b}$$

$$= \frac{2^l}{2\pi i}\oint \frac{z^2-1}{(z-x)^{n+1}}\,dz. \tag{E.26c}$$

In (E.26a), $s = \dfrac{l}{2}$, for even l,

and $s = \dfrac{l-1}{2}$, for odd l.

Also, in (E.26c), the contour of integration encloses the point $z = x$. This integral is called the *Schlaefli integral*.

$P_l(x)$ is a polynomial solution of the Legendre's differential equation,

$$(1-x^2)y'' - 2xy' + l(l+1)y = 0, \tag{E.27}$$

where $l = 0, 1, 2, \ldots + \infty$.

$P_l(x)$ is the only solution of (E.27) that is finite and analytic in the interval $-1 \le x \le 1$.

Generating Function

$$g(x,t) = (1-2xt+t^2)^{-1/2} = \sum_{l=0}^{\infty} P_l(x)t^l,\ |x| \le 1. \tag{E.28}$$

Recurrence Relations

$$(2l+1)xP_l(x) = (l+1)P_{l+1}(x) + lP_{l-1}(x), \tag{E.29a}$$

$$P'_{l+1}(x) + P'_{l-1}(x) = 2xP'_l(x) + P_l(x), \tag{E.29b}$$

$$P'_{l+1}(x) - P'_{l-1}(x) = (2l+1)P_l(x), \tag{E.29c}$$

$$(1-x^2)P'_l(x) = lP_{l-1}(x) - lxP_l(x) \tag{E.29d}$$

$$= (l+1)xP_l(x) - (l+1)P_{l+1}(x) \tag{E.29e}$$

Symmetry

$$P_l(-x) = (-1)^l P_l(x) \tag{E.30}$$

Orthogonality

$$\int_{-1}^{+1} P_l(x)P_{l'}(x)dx = \frac{2}{2l+1}\delta_{ll'}. \tag{E.31}$$

Expansion of Other Functions in Terms of $P_l(x)$

Since the $P_l(x)$ for $l = 0$ to ∞ form a complete, orthogonal set of functions, it is possible to expand any function $f(x)$ that is continuous and analytic in the interval $-1 \le x \le 1$, in terms of the Legendre polynomials:

$$f(x) = \sum_{l=0}^{\infty} a_l P_l(x), \tag{E.32}$$

where, in view of Eq. (E.31),

$$a_l = \frac{2l+1}{2}\int_{1}^{+1} f(x)P_l(x)dx. \tag{E.33}$$

The first few Legendre polynomials are listed in Table E.4.

Associated Legendre Functions

The Associated Legendre function $P_l^m(x)$ is defined by

$$P_l^m(x) = (l-x^2)^{|m|/2}\frac{d^m}{dx^m}P_l(x), \tag{E.34}$$

(where[3] $-l \le m \le l$ and, $-1 \le x \le 1$) and satisfies the differential equation,

---

3. Negative value of m is permitted in $\frac{d^m P_l}{dx^m}$, in view of the Rodrigues formula (Eq. (E.26b)) for $P_l(x)$.

Table E.4. Legendre Polynomials

l	$P_l(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2-1)$
3	$\frac{1}{2}(5x^3-3x)$
4	$\frac{1}{8}(35x^4-30x^2+3)$
5	$\frac{1}{8}(63x^5-70x^3+15x)$
	$P_l(0)=\frac{(-1)^{l/2}l!}{(l!!)^2}$, for l even
	$=0$, for l odd.

$$(1-x^2)(P_l^m)''-2x(P_l^m)'+\left\{l(l+1)-\frac{m^2}{1-x^2}\right\}P_l^m=0. \tag{E.35}$$

Generating Function

$$g(x,t)\equiv\frac{(2m-1)!!(1-x^2)^{m/2}t^m}{(1-2xt+t^2)^{m+1/2}}=\sum_{l=m}^{\infty}t^lP_l^m(x). \tag{E.36}$$

Properties

$$P_l^{-m}(x)=(-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(x) \tag{E.37}$$

$$P_l^m(-x)=(-1)^{l+m}P_l^m(x) \tag{E.38}$$

$$P_l^l(x)=(2l-1)!!(1-x^2)^{l/2}. \tag{E.39a}$$

$$P_l^0(x)=P_l(x), \tag{E.39b}$$

$$P_l^m(\pm1)=0, \text{ for } m>0, \tag{E.39c}$$

$$\int_{-1}^{+1}P_l^m(x)P_{l'}^{m'}(x)dx=\frac{2}{2l+1}\cdot\frac{(l+m)!}{(l-m)!}\delta_{ll'}\delta_{mm'}. \tag{E.40}$$

$$(2l+1)xP_l^m=(l+1-m)P_{l+1}^m+(l+m)P_{l-1}^m. \tag{E.41a}$$

$$(l-x^2)^{1/2}P_l^m=\frac{1}{2l+1}[P_{l+1}^{m+1}-P_{l-1}^{m+1}], \tag{E.41b}$$

$$= \frac{1}{2l+1}[(l+m)(l+m-1)P_{l-1}^{m-1} - (l-m+1)(l-m+2)P_{l+1}^{m-1}]. \quad \text{(E.41c)}$$

Addition Theorem for Legendre Polynomials

$$P_l(\cos\Theta) = P_l(\cos\theta_1)P_l(\cos\theta_2) + 2\sum_{m=1}^{l}\frac{(l-m)!}{l+m)!}p_l^m(\cos\theta_1)P_l^m(\cos\theta_2)$$

$$\cdot\cos\{m(\phi_1-\phi_2)\}, \quad \text{(E.42)}$$

where, the relationships between the various angles involved are indicated in Fig. 5.3.

Some of the Associated Legendre functions are listed in Table E.5.

Table E.5. Associated Legendre Functions

l	m	$P_l^m(x)$
1	1	$(1-x^2)^{1/2}$
2	1	$3x(1-x^2)^{1/2}$
	2	$3(1-x^2)$
3	1	$(3/2)(5x^2-1)(1-x^2)^{1/2}$
	2	$15x(1-x^2)$
	3	$15(1-x^2)^{3/2}$.

E.4 BESSEL FUNCTIONS

A *Bessel function of the first kind* of (integral or non-integral) order ν, is defined, for any *finite value of* x, by

$$J_\nu(x) = \sum_{s=0}^{\infty}\frac{(-1)^s}{s!\Gamma(s+\nu+1)}\left(\frac{x}{2}\right)^{\nu+2s}, \quad \text{(E.43a)}$$

$$= \frac{1}{2\pi i}\left(\frac{x}{2}\right)\oint \exp(z - x^2/4z)z^{-\nu-1}dz, \quad \text{(E.43b)}$$

$$= \frac{1}{\pi}\int_0^\pi \cos(\nu\theta - x\sin\theta)d\theta, (\nu \text{ integer}) \quad \text{(E.43c)}$$

In (E.43a), $\Gamma(\lambda)$ is the *Gamma function* given by,

$$\Gamma(\lambda) = \int_0^\infty e^{-t}t^{\lambda-1}dt. \quad \text{(E.44a)}$$

When λ is a positive integer,

$$\Gamma(\lambda) = (\lambda-1)! \quad \text{(E.44b)}$$

The contour of integration in (E.43b) is shown in Fig. E.2.

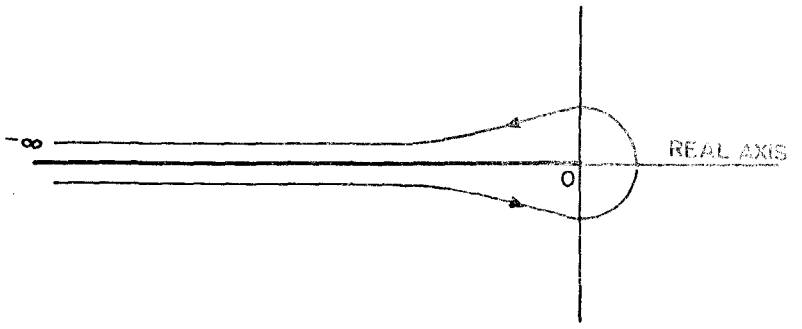


Fig. E.2. Contour of integration for the Bessel function [Eq. (E.43b)].

For $\nu = n$, where n is a positive integer,

$$J_{-n}(x) = (-1)^n J_n(x). \tag{E.45}$$

In this case, $J_n(x)$ is also called the *Bessel's coefficient*. When ν is not an integer, $J_\nu(x)$ and $J_{-\nu}(x)$ are linearly independent. $J_\nu(x)$ is a solution of the *Bessel's equation*:

$$x^2y'' + xy' + (x^2 - \nu^2)y = 0. \tag{E.46}$$

Generating Function

$$g(x,t) = \exp[(x/2)(t - 1/t)] = \sum_{n=-\infty}^{\infty} J_n(x)t^n, \ (n, \text{ integer}). \tag{E.47}$$

Recurrence Relations

$$J_{n-1}(x) + J_{n+1}(x) = (2n/x)J_n(x). \tag{E.48a}$$

$$J_{n-1}(x) - J_{n+1}(x) = 2J'_n(x) \tag{E.48b}$$

$$\frac{d}{dx}[x^n J_n(x)] = x^n J_{n-1}(x) \tag{E.48c}$$

$$\frac{d}{dx}[x^{-n} J_n(x)] = -x^{-n} J_{n+1}(x). \tag{E.48d}$$

Values for Large and Small Values of x

$$J_\nu(x) \underset{x\to\infty}{\sim} \sqrt{\frac{2}{\pi x}} \cos(x - \pi/4 - \nu\pi/2). \tag{E.49a}$$

$$J_\nu(x) \underset{x\to 0}{\sim} \frac{x^\nu}{2^\nu \Gamma(\nu+1)}. \tag{E.49b}$$

$J_\nu(x)$ is, thus, oscillatory with a decreasing amplitude for large values of x.

Orthogonality

$$\int_o^a J_\nu\left(\alpha_{\nu m}\frac{x}{a}\right)J_\nu\left(\alpha_{\nu n}\frac{x}{a}\right)x\,dx = \delta_{n,m}(a^2/2)\,[J_{\nu+1}(\alpha_{\nu m})]^2, \tag{E.50}$$

where, $\alpha_{\nu m}$ and $\alpha_{\nu n}$ are roots of the equation,

$$J_\nu(\alpha) = 0. \tag{E.51}$$

That is, $\alpha_{\nu m}$ is the mth zero of $J_\nu(\alpha)$, where $\nu > -1$. Eq. (E.50) gives the orthogonality of J_ν (for fixed ν) over the interval $0 \le x \le a$. If we further assume that the set $\{J_\nu(\alpha_{\nu m}x/a)\}$, for $m = 1, 2, 3, \ldots$ (but fixed ν), forms a complete set, then any arbitrary (but well-behaved, in the sense of Section C.1) function $f(x)$ can be expanded in terms of the Bessel functions:

$$f(x) = \sum_{m=1}^{\infty} c_{\nu m}J_\nu\left(\alpha_{\nu m}\frac{x}{a}\right),\ 0 \le x \le a;\ \nu > -1. \tag{E.52}$$

where, in view of (E.50),

$$c_{\nu m} = \frac{2}{a^2[J_{\nu+1}(\alpha_{\nu m})]^2}\int_0^a f(x)J_\nu\left(\alpha_{\nu m}\frac{x}{a}\right)x\,dx. \tag{E.53}$$

Eq. (E.52) is known as the *Bessel-Fourier Series.*

Wronskian

$$J_\nu J'_{-\nu} - J'_\nu J_{-\nu} = -\frac{2\sin\nu\pi}{\pi x}. \tag{E.54a}$$

Also

$$J_\nu J_{-\nu+1} + J_{-\nu}J_{\nu-1} = \frac{2\sin\nu\pi}{\pi x}. \tag{E.54b}$$

Neumann Functions

When ν is not an integer, $J_\nu(x)$ and $J_{-\nu}(x)$ represent the two independent solutions of the Bessel's equation. However, when ν is an integer, $J_{-\nu}$ is proportional to J_ν [(Eq. (E.45)]. In this case, the second solution of Eq. (E.46) could be chosen to be

$$N_\nu(x) = \frac{\cos\nu\pi J_\nu(x) - J_{-\nu}(x)}{\sin\nu\pi},\ (\nu \text{ integer})$$

$$= \operatorname*{Lt}_{p\to\nu}\frac{\cos p\pi J_p(x) - J_{-p}(x)}{\sin p\,\pi} \tag{E.55}$$

This is called the *Neumann function of order* ν (also referred to as the *Bessel function of the second kind*).

$N_\nu(x)$ is infinite at $x = 0$. In fact,

$$N_0(x) \underset{x\to 0}{\sim} \frac{2}{\pi}\log x, \tag{E.56a}$$

and

$$N_\nu(x) \underset{x\to 0}{\sim} -\frac{(\nu-i)!}{\pi}\left(\frac{2}{x}\right)^\nu, \nu > 0. \tag{E.56b}$$

Also,

$$N_\nu(x) \underset{x\to\infty}{\sim} \sqrt{\frac{2}{\pi x}}\sin(x-\pi/4-\nu\pi/2). \tag{E.56c}$$

Thus, for large values of x, $N_\nu(x)$ oscillates with a decreasing amplitude, just like $J_\nu(x)$ [Eq. (E.49a)].

Hankel Functions (Bessel Functions of the Third Kind)

Hankel functions are particular linear combinations of the Bessel and the Neumann functions. Thus, the Hankel functions of the first and the second kind of order ν are defined by

$$H_\nu^{(1)}(x) = J_\nu(x) + iN_\nu(x), \tag{E.57a}$$

$$H_\nu^{(2)}(x) = J_\nu(x) - iN_\nu(x). \tag{E.57b}$$

Since $J_\nu(x)$ and $N_\nu(x)$ are real, $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are complex.

Integral Representation

$$H_\nu^{(1)}(x) = \frac{1}{\pi i}\int_0^{\infty\exp(i\pi)} e^{(x/2)(t-1/t)}t^{-\nu-1}dt, \tag{E.58a}$$

$$H_\nu^{(2)}(x) = \frac{1}{\pi i}\int_{\infty\exp(i\pi)}^{0} e^{(x/2)(t-1/t)}t^{-\nu-1}dt, \tag{E.58b}$$

where the contours of integration are shown in Fig. E.3.

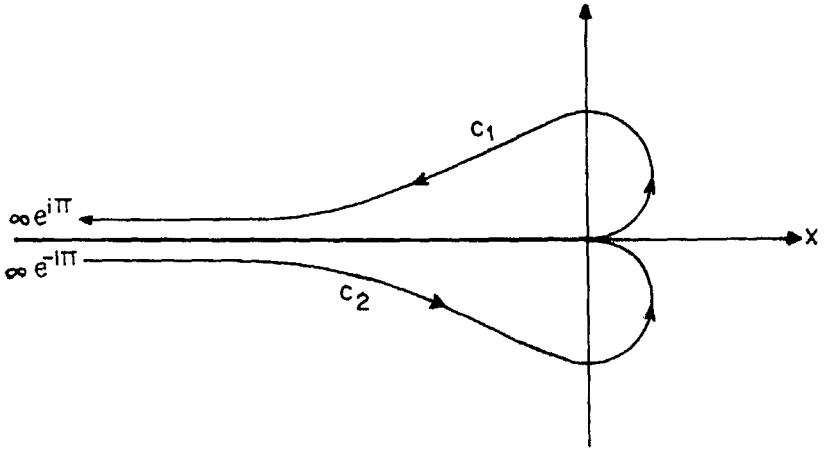


Fig. E.3. Contour of integration for the Hankel functions [Eq. (E.58a, b)].

Symmetry

$$H_{-\nu}^{(1)}(x) = e^{i\nu\pi} H_{\nu}^{(1)}(x). \tag{E.59a}$$

$$H_{-\nu}^{(2)}(x) = e^{-i\nu\pi} H_{\nu}^{(2)}(x). \tag{E.59b}$$

Modified Bessel Functions

These are solutions of the equation,

$$x^2 y''(x) + x y'(x) + (x^2 + \nu^2) y(x) = 0. \tag{E.46^1}$$

The substitution, $x = -iz$, reduces this equation to the Bessel equation (E.46) in the independent variable z. Hence the solutions of (E.46^1) are $J_\nu(ix)$ and $N_\nu(ix)$. The solutions are customarily defined as

$$I_\nu(x) = i^{-\nu} J_\nu(ix)$$

$$= e^{-i\nu\pi/2} J_\nu(x\, e^{i\pi/2}) \tag{E.60a}$$

$$= \sum_{s=0}^{\infty} \frac{1}{s!\Gamma(\nu + s + 1)} \left(\frac{x}{2}\right)^{\nu + 2s}, \tag{E.60b}$$

and

$$K_\nu(x) = \frac{\pi}{2} i^{\nu+1} [J_\nu(ix) + i N_\nu(ix)]$$

$$= (\pi/2) \cdot \frac{I_{-\nu}(x) - I_\nu(x)}{\sin \nu\pi}. \tag{E.61}$$

$I_\nu(x)$ and $K_\nu(x)$ are, respectively, the *Modified Bessel Functions* of the First and the Second Kind. The latter is also called the *Basset Function.*

Unlike the Bessel and the Neumann Functions, the functions $I_\nu(x)$ and $K_\nu(x)$ are not oscillatory, but are exponential for large values of x:

$$I_0(x) \underset{x \to \infty}{\sim} \frac{1}{\sqrt{2\pi x}} e^x, \tag{E.62a}$$

$$K_0(x) \underset{x \to \infty}{\sim} \sqrt{\frac{\pi}{2x}} e^{-x}. \tag{E.62b}$$

Also,

$$I_0(x) \underset{x \to 0}{\sim} 1, \tag{E.63a}$$

$$K_0(x) \underset{x \to 0}{\sim} -\ln (x/2). \tag{E.63b}$$

Spherical Bessel Functions

In the equation,

$$\frac{d^2R}{dr^2}+\frac{2}{r}\frac{dR}{dr}+\left\{k^2-\frac{l(l+1)}{r^2}\right\}R=0, \tag{E.64}$$

where $l=0,1,2,\ldots,+\infty$, if we make the substitution,

$$\rho=kr;\ y(\rho)=\sqrt{\rho}R(r), \tag{E.65}$$

we get

$$\rho^2y''+\rho y'+\left\{\rho^2-\left(l+\frac{1}{2}\right)^2\right\}y=0,$$

which is Eq. (E.46) with $\nu=l+\frac{1}{2}$. Hence, the two independent solutions of Eq. (E.64) are proportional to $(1/\sqrt{\rho})J_{l+1/2}(\rho)$ and $\frac{1}{\sqrt{\rho}}J_{-l-1/2}(\rho)$. These are usually defined as

$$j_l(\rho)=\sqrt{\frac{\pi}{2\rho}}J_{l+1/2}(\rho)=(2\rho)^l\sum_{s=0}^{\infty}\left\{\frac{(-1)^s(s+l)!}{s!(2s+2l+1)!}\right\}\rho^{2s}, \tag{E.66a}$$

$$n_l(\rho)=\sqrt{\frac{\pi}{2\rho}}N_{l+1/2}(\rho)=(-1)^{l+1}\sqrt{\frac{\pi}{2\rho}}J_{-l-1/2}(\rho)$$

$$=-\frac{1}{2^l\rho^{l+1}}\sum_{s=0}^{\infty}\left\{\frac{(2l-2s)!}{s!(l-s)!}\right\}\rho^{2s}. \tag{E.66b}$$

(E.66a) and (E.66b) are, respectively called the *spherical Bessel* and the *spherical Neumann* Functions. In analogy with Eqs. (E.57a, b), the *spherical Hankel Functions* of the first and the second kind are defined by,

$$h_l^{(1)}(\rho)=j_l(\rho)+in_l(\rho)=-ih_l^{(+)}(\rho), \tag{E.67a}$$

$$h_l^{(2)}(\rho)=j_l(\rho)-in_l(\rho)=+ih_l^{(-)}(\rho). \tag{E.67b}$$

From Eqs. (E.66a, b), we have,

$$j_0(\rho)=\frac{\sin\rho}{\rho}, \tag{E.68a}$$

$$n_0(\rho)=-\frac{\cos\rho}{\rho}, \tag{E.68b}$$

$$h_0^{\pm}(\rho)=\frac{e^{\pm i\rho}}{\rho}. \tag{E.68c}$$

Also,

$$j_l(\rho)\underset{\rho\to 0}{\sim}\frac{\rho^l}{(2l+1)!!}, \tag{E.69a}$$

$$n_l(\rho) \underset{\rho \to 0}{\sim} -(2l-1)!!\rho^{-l-1}. \tag{E.69b}$$

$$j_l(\rho) \underset{\rho \to \infty}{\sim} \frac{1}{\rho}\sin\left(\rho - \frac{l\pi}{2}\right), \tag{E.70a}$$

$$n_l(\rho) \underset{\rho \to \infty}{\sim} -\frac{1}{\rho}\cos\left(\rho - \frac{l\pi}{2}\right), \tag{E.70b}$$

$$h_l^{(\pm)}(\rho) \underset{\rho \to \infty}{\sim} \frac{1}{\rho} e^{\pm i(\rho - l\pi/2)} \tag{E.70c}$$

Recurrence Relations ($l \geq 1$)

$$f_{l-1}(\rho) + f_{l+1}(\rho) = \frac{2l+1}{\rho} f_l(\rho), \tag{E.71a}$$

$$l f_{l-1}(\rho) - (l+1) f_{l+1}(\rho) = (2l+1) f'_l(\rho), \tag{E.71b}$$

$$\frac{d}{d\rho}[\rho^{l-1} f_l(\rho)] = \rho^{l+1} f_{l-1}(\rho), \tag{E.71c}$$

$$\frac{d}{d\rho}[\rho^{-l} f_l(\rho)] = -\rho^{-l} f_{l+1}(\rho), \tag{E.71d}$$

where $f_l(\rho)$ stands for $j_l, n_l, h_l^{(1)}, h_l^{(2)}$ or $h_l^{(\pm)}$.

Rayleigh and Other Formulae

$$j_l(\rho) = (-1)^l \rho^l \left(\frac{d}{\rho d\rho}\right)^l \left(\frac{\sin\rho}{\rho}\right), \tag{E.72a}$$

$$n_l(\rho) = (-1)^{l+1} \rho^l \left(\frac{d}{\rho d\rho}\right)^l \left(\frac{\cos\rho}{\rho}\right), \tag{E.72b}$$

$$h_l^{(\pm)}(\rho) = (-1)^l \rho^l \left(\frac{d}{d\rho}\right)^l \left(\frac{e^{\pm\rho}}{\rho}\right). \tag{E.72c}$$

$$j_l(\rho) n'_l(\rho) - j'_l(\rho) n_l(\rho) = \frac{1}{\rho^2}. \tag{E.73}$$

Orthogonality

$$\int_{-\infty}^{+\infty} j_l(x) j_{l'}(x) dx = \delta_{ll'} \left(\frac{\pi}{2l+1}\right). \tag{E.74}$$

This can be deduced from Eq. (E.50).

Addition Theorem

$$\sum_{l=0}^{\infty} (2l+1)\{j_l(kr)\}^2 P_l(\cos\theta) = \frac{\sin Kr}{Kr}, \tag{E.75}$$

where

$$K = 2k \sin(\theta/2).$$

$$\frac{\exp(ik\,|\,\mathbf{r}-\mathbf{r}'\,|)}{|\,\mathbf{r}-\mathbf{r}'\,|} = ik \sum_{l=0}^{\infty} (2l+1) j_l(kr_<) h_l^{(1)}(kr_>) P_l(\cos\theta), \tag{E.76}$$

θ being the angle between $\mathbf{r}$ and $\mathbf{r}'$.

THIS PAGE IS
BLANK

INDEX

THIS PAGE IS
BLANK