The Physics of Quantum Mechanics

James Binney
and
David Skinner

Published by Capella Archive 2008

Contents

Appendices

Preface

This book grew out of classes given for many years to the second-year undergraduates of Merton College, Oxford. The University lectures that the students were attending in parallel were restricted to the wave-mechanical methods introduced by Schrödinger, with a very strong emphasis on the time-independent Schrödinger equation. The classes had two main aims: to introduce more wide-ranging concepts associated especially with Dirac and Feynman, and to give the students a better understanding of the physical implications of quantum mechanics as a description of how systems great and small evolve in time.

While it is important to stress the revolutionary aspects of quantum mechanics, it is no less important to understand that classical mechanics is just an approximation to quantum mechanics. Traditional introductions to quantum mechanics tend to neglect this task and leave students with two independent worlds, classical and quantum. At every stage we try to explain how classical physics emerges from quantum results. This exercise helps students to extend to the quantum regime the intuitive understanding they have developed in the classical world. This extension both takes much of the mystery from quantum results, and enables students to check their results for common sense and consistency with what they already know.

A key to understanding the quantum–classical connection is the study of the evolution in time of quantum systems. Traditional texts stress instead the recovery of stationary states, which do not evolve. We want students to understand that the world is full of change – that dynamics *exists* – precisely because the energies of real systems are always uncertain, so a real system is never in a stationary state; stationary states are useful mathematical abstractions but are not physically realisable. We try to avoid confusion between the real physical novelty in quantum mechanics and the particular way in which it is convenient to solve its governing equation, the time-dependent Schrödinger equation.

Quantum mechanics emerged from efforts to understand atoms, so it is natural that atomic physics looms large in traditional courses. However, atoms are complex systems in which tens of particles interact strongly with each other at relativistic speeds. We believe it is a mistake to plunge too soon into this complex field. We cover atoms only in so far as we can proceed with a reasonable degree of rigour. This includes hydrogen and helium in some detail (including a proper treatment of Thomas precession), and a qualitative sketch of the periodic table. But is excludes traditional topics such as spin–orbit coupling schemes and the physical interpretation of atomic spectra.

We devote a chapter to the adiabatic principle, which opens up a wonderfully rich range of phenomena to quantitative investigation. We also devote a chapter to scattering theory, which is both an important practical application of quantum mechanics, and a field that raises some interesting conceptual issues and makes one think carefully about how we compute results in quantum mechanics.

When one sits down to solve a problem in physics, it's vital to identify the optimum coordinate system for the job – a problem that is intractable in the coordinate system that first comes to mind, may be trivial in another system. Dirac's notation makes it possible to think about physical problems in a coordinate-free way, and makes it straightforward to move to the chosen coordinate system once that has been identified. Moreover, Dirac's notation brings into sharp focus the still mysterious concept of a probability amplitude. Hence, it is important to introduce Dirac's notation from the outset, and to use it for an extensive discussion of probability amplitudes and why they lead to qualitatively new phenomena.

In the winter of 2008/9 the book was used as the basis for the second-year introductory quantum-mechanics course in Oxford Physics. At the outset there was a whiff of panic in the air, emanating from tutors as well as students. Gradually more and more participants grasped what was going on and appreciated the intellectual excitement of the subject. Although the final feedback covered the full gamut of opinion from "incomprehensible" to "the best course ever" there were clear indications that many students and some tutors had risen to the challenge and gained a deeper understanding of this difficult subject than was previously usual. Several changes to the text of this revised edition were made in response to feedback from students and tutors. It was clear that students needed to be given more time to come to terms with quantum amplitudes and Dirac notation. To this end some work on spin-half systems and polarised light has been introduced to Chapter 1. The students found orbital angular momentum hard, and the way this is handled in what is now Chapter 7 has been changed.

The major change from the first edition is Chapter 6, a new chapter on composite systems. It starts with material transferred from the end of Chapter 2 of the first edition, and then discusses entanglement, the Einstein–Podolski–Rosen experiment and Bell inequalities. Sections on quantum computing, density operators, thermodynamics and the measurement problem follow. It is most unusual for the sixth chapter of a second-year textbook to be able to take students to the frontier of human understanding, as this chapter does. More minor changes include the addition of a section on the Heisenberg picture to Chapter 4, the correction of a widespread misunderstanding about the singlet-triplet splitting in helium in Chapter 10, and the addition of thermodynamics to the applications of the adiabatic principle discussed in Chapter 11.

Problem solving is the key to learning physics and most chapters are followed by a long list of problems. These lists have been extensively revised since the first edition and printed solutions prepared. The solutions to starred problems, which are mostly more-challenging problems, are now available online[1] and solutions to other problems are available to colleagues who are teaching a course from the book. In nearly every problem a student will either prove a useful result or deepen his/her understanding of quantum mechanics and what it says about the material world. Even after successfully solving a problem we suspect students will find it instructive and thought-provoking to study the solution posted on the web.

We are grateful to several colleagues for comments on the first edition, particularly Justin Wark for alerting us to the problem with the singlet-triplet splitting. Fabian Essler and John March-Russell made several constructive suggestions. We thank our fellow Mertonian Artur Ekert for stimulating discussions of material covered in Chapter 6 and for reading that chapter in draft form.

September 2009

James Binney
David Skinner

---

[1] http://www-thphys.physics.ox.ac.uk/users/JamesBinney/QBhome.htm

1

Probability and probability amplitudes

The future is always uncertain. Will it rain tomorrow? Will Pretty Lady win the 4.20 race at Sandown Park on Tuesday? Will the Financial Times All Shares index rise by more than 50 points in the next two months? Nobody knows the answers to such questions, but in each case we may have information that makes a positive answer more or less appropriate: if we are in the Great Australian Desert and it's winter, it is exceedingly unlikely to rain tomorrow, but if we are in Delhi in the middle of the monsoon, it will almost certainly rain. If Pretty Lady is getting on in years and hasn't won a race yet, she's unlikely to win on Tuesday either, while if she recently won a couple of major races and she's looking fit, she may well win at Sandown Park. The performance of the All Shares index is hard to predict, but factors affecting company profitability and the direction interest rates will move, will make the index more or less likely to rise. Probability is a concept which enables us to quantify and manipulate uncertainties. We assign a probability $p = 0$ to an event if we think it is simply impossible, and we assign $p = 1$ if we think the event is certain to happen. Intermediate values for p imply that we think an event may happen and may not, the value of p increasing with our confidence that it will happen.

Physics is about predicting the future. Will this ladder slip when I step on it? How many times will this pendulum swing to and fro in an hour? What temperature will the water in this thermos be at when it has completely melted this ice cube? Physics often enables us to answer such questions with a satisfying degree of certainty: the ladder will not slip provided it is inclined at less than 23.34° to the vertical; the pendulum makes 3602 oscillations per hour; the water will reach 6.43°C. But if we are pressed for sufficient accuracy we must admit to uncertainty and resort to probability because our predictions depend on the data we have, and these are always subject to measuring error, and idealisations: the ladder's critical angle depends on the coefficients of friction at the two ends of the ladder, and these cannot be precisely given because both the wall and the floor are slightly irregular surfaces; the period of the pendulum depends slightly on the amplitude of its swing, which will vary with temperature and the humidity of the air; the final temperature of the water will vary with the amount of heat transferred through the walls of the thermos and the speed of evaporation from the water's surface, which depends on draughts in the room as well as on humidity. If we are asked to make predictions about a ladder that is inclined near its critical angle, or we need to know a quantity like the period of the pendulum to high accuracy, we cannot make definite statements, we can only say something like the probability of the ladder slipping is 0.8, or there is a probability of 0.5 that the period of the pendulum lies between 1.0007 s and 1.0004 s. We can dispense with probability when slightly vague answers are permissible, such as that the period is 1.00 s to three significant figures. The concept of probability enables us to push our science to its limits, and make the most precise and reliable statements possible.

Probability enters physics in two ways: through uncertain data and through the system being subject to random influences. In the first case we could make a more accurate prediction if a property of the system, such as the length or temperature of the pendulum, were more precisely characterised. That is, the value of some number is well defined, it's just that we don't know the value very accurately. The second case is that in which our system is subject to inherently random influences – for example, to the draughts that make us uncertain what will be the final temperature of the water. To attain greater certainty when the system under study is subject to such random influences, we can either take steps to increase the isolation of our system – for example by putting a lid on the thermos – or we can expand the system under study so that the formerly random influences become calculable interactions between one part of the system and another. Such expansion of the system is not a practical proposition in the case of the thermos – the expanded system would have to encompass the air in the room, and then we would worry about fluctuations in the intensity of sunlight through the window, draughts under the door and much else. The strategy does work in other cases, however. For example, climate changes over the last ten million years can be studied as the response of a complex dynamical system – the atmosphere coupled to the oceans – that is subject to random external stimuli, but a more complete account of climate changes can be made when the dynamical system is expanded to include the Sun and Moon because climate is strongly affected by the inclination of the Earth's spin axis to the plane of the Earth's orbit and the luminosity of the Sun.

A low-mass system is less likely to be well isolated from its surroundings than a massive one. For example, the orbit of the Earth is scarcely affected by radiation pressure that sunlight exerts on it, while dust grains less than a few microns in size that are in orbit about the Sun lose angular momentum through radiation pressure at a rate that causes them to spiral in from near the Earth to the Sun within a few millennia. Similarly, a rubber duck left in the bath after the children have got out will stay very still, while tiny pollen grains in the water near it execute Brownian motion that carries them along a jerky path many times their own length each minute. Given the difficulty of isolating low-mass systems, and the tremendous obstacles that have to be surmounted if we are to expand the system to the point at which all influences on the object of interest become causal, it is natural that the physics of small systems is invariably probabilistic in nature. Quantum mechanics describes the dynamics of all systems, great and small. Rather than making firm predictions, it enables us to calculate probabilities. If the system is massive, the probabilities of interest may be so near zero or unity that we have effective certainty. If the system is small, the probabilistic aspect of the theory will be more evident.

The scale of atoms is precisely the scale on which the probabilistic aspect is predominant. Its predominance reflects two facts. First, there is no such thing as an isolated atom because all atoms are inherently coupled to the electromagnetic field, and to the fields associated with electrons, neutrinos, quarks, and various 'gauge bosons'. Since we have incomplete information about the states of these fields, we cannot hope to make precise predictions about the behaviour of an individual atom. Second, we cannot build measuring instruments of arbitrary delicacy. The instruments we use to measure atoms are usually themselves made of atoms, and employ electrons or photons that carry sufficient energy to change an atom significantly. We rarely know the exact state that our measuring instrument is in before we bring it into contact with the system we have measured, so the result of the measurement of the atom would be uncertain even if we knew the precise state that the atom was in before we measured it, which of course we do not. Moreover, the act of measurement inevitably disturbs the atom, and leaves it in a different state from the one it was in before we made the measurement. On account of the uncertainty inherent in the measuring process, we cannot be sure what this final state may be. Quantum mechanics allows us to calculate probabilities for each possible final state. Perhaps surprisingly, from the theory it emerges that even when we have the most complete information about the state of a system that is is logically possible to have, the outcomes of some measurements remain uncertain. Thus whereas in the classical world uncertainties can be made as small as we please by sufficiently careful work, in the quantum world uncertainty is woven into the fabric of reality.

1.1 The laws of probability

Events are frequently one-offs: Pretty Lady will run in the 4.20 at Sandown Park only once this year, and if she enters the race next year, her form and the field will be different. The probability that we want is for this year's race. Sometimes events can be repeated, however. For example, there is no obvious difference between one throw of a die and the next throw, so it makes sense to assume that the probability of throwing a 5 is the same on each throw. When events can be repeated in this way we seek to assign probabilities in such a way that when we make a very large number N of trials, the number n_A of trials in which event A occurs (for example 5 comes up) satisfies

$$n_A \simeq p_A N. \tag{1.1}$$

In any realistic sequence of throws, the ratio n_A/N will vary with N, while the probability p_A does not. So the relation (1.1) is rarely an equality. The idea is that we should choose p_A so that n_A/N fluctuates in a smaller and smaller interval around p_A as N is increased.

Events can be logically combined to form composite events: if A is the event that a certain red die falls with 1 up, and B is the event that a white die falls with 5 up, AB is the event that when both dice are thrown, the red die shows 1 and the white one shows 5. If the probability of A is p_A and the probability of B is p_B, then in a fraction $\sim p_A$ of throws of the two dice the red die will show 1, and in a fraction $\sim p_B$ of these throws, the white die will have 5 up. Hence the fraction of throws in which the event AB occurs is $\sim p_A p_B$ so we should take the probability of AB to be $p_{AB} = p_A p_B$. In this example A and B are **independent** events because we see no reason why the number shown by the white die could be influenced by the number that happens to come up on the red one, and vice versa. The rule for combining the probabilities of independent events to get the probability of both events happening, is to multiply them:

$$p(A \text{ and } B) = p(A)p(B) \quad \text{(independent events)}. \tag{1.2}$$

Since only one number can come up on a die in a given throw, the event A above excludes the event C that the red die shows 2; A and C are **exclusive** events. The probability that *either* a 1 or a 2 will show is obtained by adding p_A and p_C. Thus

$$p(A \text{ or } C) = p(A) + p(C) \quad \text{(exclusive events)}. \tag{1.3}$$

In the case of reproducible events, this rule is clearly consistent with the principle that the fraction of trials in which either A or C occurs should be the sum of the fractions of the trials in which one or the other occurs. If we throw our die, the number that will come up is certainly one of 1, 2, 3, 4, 5 or 6. So by the rule just given, the sum of the probabilities associated with each of these numbers coming up has to be unity. Unless we know that the die is loaded, we assume that no number is more likely to come up than another, so all six probabilities must be equal. Hence, they must all equal $\frac{1}{6}$. Generalising this example we have the rules

$$\begin{aligned}&\text{With just } N \text{ mutually exclusive outcomes, } \sum_{i=1}^{N} p_i = 1.\\ &\text{If all outcomes are equally likely, } p_i = 1/N.\end{aligned} \tag{1.4}$$

1.1.1 Expectation values

A **random variable** x is a quantity that we can measure and the value that we get is subject to uncertainty. Suppose for simplicity that only discrete values x_i can be measured. In the case of a die, for example, x could be the number that comes up, so x has six possible values, $x_1 = 1$ to $x_6 = 6$. If p_i is the probability that we shall measure x_i, then the **expectation value** of x is

$$\langle x\rangle \equiv \sum_i p_i x_i. \tag{1.5}$$

If the event is reproducible, it is easy to show that the average of the values that we measure on N trials tends to $\langle x\rangle$ as N becomes very large. Consequently, $\langle x\rangle$ is often referred to as the average of x.

Suppose we have two random variables, x and y. Let p_{ij} be the probability that our measurement returns x_i for the value of x and y_j for the value of y. Then the expectation of the sum $x + y$ is

$$\langle x+y\rangle = \sum_{ij} p_{ij}(x_i + y_j) = \sum_{ij} p_{ij} x_i + \sum_{ij} p_{ij} y_j \tag{1.6}$$

But $\sum_j p_{ij}$ is the probability that we measure x_i regardless of what we measure for y, so it must equal p_i. Similarly $\sum_i p_{ij} = p_j$, the probability of measuring y_j irrespective of what we get for x. Inserting these expressions in to (1.6) we find

$$\langle x+y\rangle = \langle x\rangle + \langle y\rangle\,. \tag{1.7}$$

That is, the expectation value of the sum of two random variables is the sum of the variables' individual expectation values, regardless of whether the variables are independent or not.

A useful measure of the amount by which the value of a random variable fluctuates from trial to trial is the **variance** of x:

$$\left\langle (x - \langle x\rangle)^2\right\rangle = \left\langle x^2\right\rangle - 2\left\langle x\left\langle x\right\rangle\right\rangle + \left\langle \left\langle x\right\rangle^2\right\rangle, \tag{1.8}$$

where we have made use of equation (1.7). The expectation $\langle x\rangle$ is not a random variable, but has a definite value. Consequently $\langle x\langle x\rangle\rangle = \langle x\rangle^2$ and $\left\langle \langle x\rangle^2\right\rangle = \langle x\rangle^2$, so the variance of x is related to the expectations of x and x^2 by

$$\left\langle \Delta_x^2\right\rangle \equiv \left\langle (x - \langle x\rangle)^2\right\rangle = \left\langle x^2\right\rangle - \langle x\rangle^2\,. \tag{1.9}$$

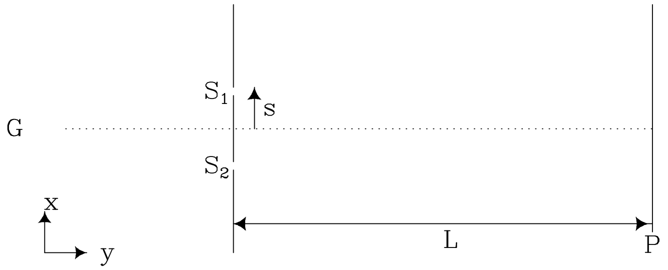


Figure 1.1 The two-slit interference experiment.

1.2 Probability amplitudes

Many branches of the social, physical and medical sciences make extensive use of probabilities, but quantum mechanics stands alone in the *way* that it calculates probabilities, for it always evaluates a probability p as the mod-square of a certain complex number A:

$$p = |A|^2. \tag{1.10}$$

The complex number A is called the **probability amplitude** for p.

Quantum mechanics is the only branch of knowledge in which probability amplitudes appear, and nobody understands why they arise. They give rise to phenomena that have no analogues in classical physics through the following fundamental principle. Suppose something can happen by two (mutually exclusive) routes, S or T, and let the probability amplitude for it to happen by route S be $A(S)$ and the probability amplitude for it to happen by route T be $A(T)$. Then the probability amplitude for it to happen by one route or the other is

$$A(S \text{ or } T) = A(S) + A(T). \tag{1.11}$$

This rule takes the place of the sum rule for probabilities, equation (1.3). However, it is incompatible with equation (1.3), because it implies that the probability that the event happens regardless of route is

$$\begin{aligned} p(S \text{ or } T) &= |A(S \text{ or } T)|^2 = |A(S) + A(T)|^2 \\ &= |A(S)|^2 + A(S)A^*(T) + A^*(S)A(T) + |A(T)|^2 \\ &= p(S) + p(T) + 2\Re\mathrm{e}(A(S)A^*(T)). \end{aligned} \tag{1.12}$$

That is, the probability that an event will happen is not merely the sum of the probabilities that it will happen by each of the two possible routes: there is an additional term $2\Re\mathrm{e}(A(S)A^*(T))$. This term has no counterpart in standard probability theory, and violates the fundamental rule (1.3) of probability theory. It depends on the phases of the probability amplitudes for the individual routes, which do not contribute to the probabilities $p(S) = |A(S)|^2$ of the routes.

Whenever the probability of an event differs from the sum of the probabilities associated with the various mutually exclusive routes by which it can happen, we say we have a manifestation of **quantum interference**. The term $2\Re\mathrm{e}(A(S)A^*(T))$ in equation (1.12) is what generates quantum interference mathematically. We shall see that in certain circumstances the violations of equation (1.3) that are caused by quantum interference are not detectable, so standard probability theory appears to be valid.

How do we know that the principle (1.11), which has these extraordinary consequences, is true? The soundest answer is that it is a fundamental postulate of quantum mechanics, and that every time you look at a digital watch, or touch a computer keyboard, or listen to a CD player, or interact with any other electronic device that has been engineered with the help of quantum mechanics, you are testing and vindicating this theory. Our civilisation now quite simply depends on the validity of equation (1.11).

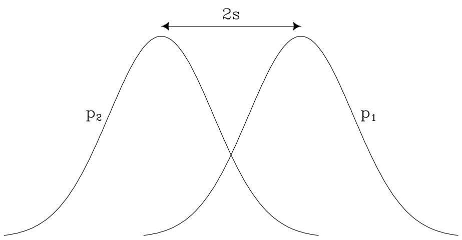


Figure 1.2 The probability distributions of passing through each of the two closely spaced slits overlap.

1.2.1 Two-slit interference

An imaginary experiment will clarify the physical implications of the principle and suggest how it might be tested experimentally. The apparatus consists of an electron gun, G, a screen with two narrow slits S_1 and S_2, and a photographic plate P, which darkens when hit by an electron (see Figure 1.1).

When an electron is emitted by G, it has an amplitude to pass through slit S_1 and then hit the screen at the point x. This amplitude will clearly depend on the point x, so we label it $A_1(x)$. Similarly, there is an amplitude $A_2(x)$ that the electron passed through S_2 before reaching the screen at x. Hence the probability that the electron arrives at x is

$$P(x) = |A_1(x) + A_2(x)|^2 = |A_1(x)|^2 + |A_2(x)|^2 + 2\Re\mathrm{e}(A_1(x)A_2^*(x)). \quad (1.13)$$

$|A_1(x)|^2$ is simply the probability that the electron reaches the plate after passing through S_1. We expect this to be a roughly Gaussian distribution $p_1(x)$ that is centred on the value x_1 of x at which a straight line from G through the middle of S_1 hits the plate. $|A_2(x)|^2$ should similarly be a roughly Gaussian function $p_2(x)$ centred on the intersection at x_2 of the screen and the straight line from G through the middle of S_2. It is convenient to write $A_i = |A_i|\mathrm{e}^{\mathrm{i}\phi_i} = \sqrt{p_i}\mathrm{e}^{\mathrm{i}\phi_i}$, where ϕ_i is the phase of the complex number A_i. Then equation (1.13) can be written

$$p(x) = p_1(x) + p_2(x) + I(x), \quad (1.14\text{a})$$

where the **interference term** I is

$$I(x) = 2\sqrt{p_1(x)p_2(x)}\cos(\phi_1(x) - \phi_2(x)). \quad (1.14\text{b})$$

Consider the behaviour of $I(x)$ near the point that is equidistant from the slits. Then (see Figure 1.2) $p_1 \simeq p_2$ and the interference term is comparable in magnitude to $p_1 + p_2$, and, by equations (1.14), the probability of an electron arriving at x will oscillate between $\sim 2p_1$ and 0 depending on the value of the phase difference $\phi_1(x) - \phi_2(x)$. In §2.3.4 we shall show that the phases $\phi_i(x)$ are approximately linear functions of x, so after many electrons have been fired from G to P in succession, the blackening of P at x, which will be roughly proportional to the number of electrons that have arrived at x, will show a sinusoidal pattern.

Let's replace the electrons by machine-gun bullets. Then everyday experience tells us that classical physics applies, and it predicts that the probability $p(x)$ of a bullet arriving at x is just the sum $p_1(x) + p_2(x)$ of the probabilities of a bullet coming through S_1 or S_2. Hence classical physics does not predict a sinusoidal pattern in $p(x)$. How do we reconcile the very different predictions of classical and quantum mechanics? Firearms manufacturers have for centuries used classical mechanics with deadly success, so is the resolution that bullets do not obey quantum mechanics? We believe they do, and the probability distribution for the arrival of bullets *should* show a sinusoidal pattern. However, in §2.3.4 we shall find that quantum mechanics predicts that the distance Δ between the peaks and troughs of this pattern becomes smaller and smaller as we increase the mass of the particles we are firing through the slits, and by the time the particles are as massive as a bullet, Δ is fantastically small. Consequently, it is not experimentally feasible to test whether $p(x)$ becomes small at regular intervals. Any feasible experiment will probe the value of $p(x)$ averaged over many peaks and troughs of the sinusoidal pattern. This averaged value of $p(x)$ agrees with the probability distribution we derive from classical mechanics because the average value of $I(x)$ in equation (1.14) vanishes.

1.2.2 Matter waves?

The sinusoidal pattern of blackening on P that quantum mechanics predicts proves to be identical to the interference pattern that is observed in Young's double-slit experiment. This experiment established that light is a wave phenomenon because the wave theory could readily explain the existence of the interference pattern. It is natural to infer from the existence of the sinusoidal pattern in the quantum-mechanical case, that particles are manifestations of waves in some medium. There is much truth in this inference, and at an advanced level this idea is embodied in quantum field theory. However, in the present context of non-relativistic quantum mechanics, the concept of matter waves is unhelpful. Particles are particles, not waves, and they pass through one slit or the other. The sinusoidal pattern arises because probability amplitudes are complex numbers, which add in the same way as wave amplitudes. Moreover, the energy density (intensity) associated with a wave is proportional to the mod square of the wave amplitude, just as the probability density of finding a particle is proportional to the mod square of the probability amplitude. Hence, on a mathematical level, there is a one-to-one correspondence between what happens when particles are fired towards a pair of slits and when light diffracts through similar slits. But we cannot consistently infer from this correspondence that particles are manifestations of waves because quantum interference occurs in quantum systems that are much more complex than a single particle, and indeed in contexts where motion through space plays no role. In such contexts we cannot ascribe the interference phenomenon to interference between real physical waves, so it is inconsistent to take this step in the case of single-particle mechanics.

1.3 Quantum states

1.3.1 Quantum amplitudes and measurements

Physics is about the quantitative description of natural phenomena. A quantitative description of a system inevitably starts by defining ways in which it can be measured. If the system is a single particle, quantities that we can measure are its x, y and z coordinates with respect to some choice of axes, and the components of its momentum parallel to these axes. We can also measure its energy, and its angular momentum. The more complex a system is, the more ways there will be in which we can measure it.

Associated with every measurement, there will be a set of possible numerical values for the measurement – the **spectrum** of the measurement. For example, the spectrum of the x coordinate of a particle in empty space is the interval $(-\infty, \infty)$, while the spectrum of its kinetic energy is $(0, \infty)$. We shall encounter cases in which the spectrum of a measurement consists of discrete values. For example, in Chapter 7 we shall show that the angular momentum of a particle parallel to any given axis has spectrum $(\ldots, (k-1)\hbar, k\hbar, (k+1)\hbar, \ldots)$, where $\hbar$ is Planck's constant $h = 6.63 \times 10^{-34}\,\mathrm{J\,s}$ divided by 2π, and k is either 0 or $\frac{1}{2}$. When the spectrum is a set of discrete numbers, we say that those numbers are the **allowed** values of the measurement.

With every value in the spectrum of a given measurement there will be a quantum amplitude that we will find this value if we make the relevant measurement. Quantum mechanics is the science of how to calculate such amplitudes given the results of a sufficient number of prior measurements.

Imagine that you're investigating some physical system: some particles in an ion trap, a drop of liquid helium, the electromagnetic field in a resonant cavity. What do you know about the state of this system? You have two types of knowledge: (1) a specification of the physical nature of the system (e.g., size & shape of the resonant cavity), and (2) information about the current dynamical state of the system. In quantum mechanics information of type (1) is used to define an object called the Hamiltonian H of the system that is defined by equation (2.5) below. Information of type (2) is more subtle. It must consist of predictions for the outcomes of measurements you could make on the system. Since these outcomes are inherently uncertain, your information must relate to the probabilities of different outcomes, and in the simplest case consists of values for the relevant probability amplitudes. For example, your knowledge might consist of amplitudes for the various possible outcomes of a measurement of energy, or of a measurement of momentum.

In quantum mechanics, then, knowledge about the current dynamical state of a system is embodied in a set of quantum amplitudes. In classical physics, by contrast, we can state with certainty which value we will measure, and we characterise the system's current dynamical state by simply giving this value. Such values are often called 'coordinates' of the system. Thus in quantum mechanics a whole set of quantum amplitudes replaces a single number.

Complete sets of amplitudes Given the amplitudes for a certain set of events, it is often possible to calculate amplitudes for other events. The phenomenon of particle spin provides the neatest illustration of this statement.

Electrons, protons, neutrinos, quarks, and many other elementary particles turn out to be tiny gyroscopes: they spin. The rate at which they spin and therefore the the magnitude of their spin angular momentum never changes; it is always $\sqrt{3/4}\hbar$. Particles with this amount of spin are called **spin-half particles** for reasons that will emerge shortly. Although the spin of a spin-half particle is fixed in magnitude, its direction can change. Consequently, the value of the spin angular momentum parallel to any given axis can take different values. In §7.4.2 we shall show that parallel to any given axis, the spin angular momentum of a spin-half particle can be either $\pm\frac{1}{2}\hbar$. Consequently, the spin parallel to the z axis is denoted $s_z\hbar$, where $s_z = \pm\frac{1}{2}$ is an observable with the spectrum $\{-\frac{1}{2}, \frac{1}{2}\}$.

In §7.4.2 we shall show that if we know *both* the amplitude a_+ that s_z will be measured to be $+\frac{1}{2}$ *and* the amplitude a_- that a measurement will yield $s_z = -\frac{1}{2}$, then we can calculate from these two complex numbers the amplitudes b_+ and b_- for the two possible outcomes of the measurement of the spin along *any* direction. If we know only a_+ (or only a_-), then we can calculate neither b_+ nor b_- for *any* other direction.

Generalising from this example, we have the concept of a **complete set of amplitudes**: the set contains enough information to enable one to calculate amplitudes for the outcome of *any* measurement whatsoever. Hence, such a set gives a complete specification of the physical state of the system. A complete set of amplitudes is generally understood to be a minimal set in the sense that none of the amplitudes can be calculated from the others. The set $\{a_-, a_+\}$ consitutes a complete set of amplitudes for the spin of an electron.

1.3.2 Dirac notation

Dirac introduced the symbol $|\psi\rangle$, pronounced 'ket psi', to denote a complete set of amplitudes for the system. If the system consists of a particle[1] trapped in a potential well, $|\psi\rangle$ could consist of the amplitudes a_n that the energy is E_n, where $(E_1, E_2, \ldots)$ is the spectrum of possible energies, or it might consist of the amplitudes $\psi(x)$ that the particle is found at x, or it might consist of the amplitudes $a(p)$ that the momentum is measured to be p. Using the abstract symbol $|\psi\rangle$ enables us to think about the system without committing ourselves to what complete set of amplitudes we are going to use, in the same way that the position vector $\mathbf{x}$ enables us to think about a geometrical point independently of the coordinates (x, y, z), (r, θ, ϕ) or whatever by which we locate it. That is, $|\psi\rangle$ is a container for a complete set of amplitudes in the same way that a vector $\mathbf{x}$ is a container for a complete set of coordinates.

The ket $|\psi\rangle$ encapsulates the crucial concept of a **quantum state**, which is independent of the particular set of amplitudes that we choose to quantify it, and is fundamental to several branches of physics.

We saw in the last section that amplitudes must sometimes be added: if an outcome can be achieved by two different routes and we do not monitor the route by which it is achieved, we add the amplitudes associated with each route to get the overall amplitude for the outcome. In view of this additivity, we write

$$|\psi_3\rangle = |\psi_1\rangle + |\psi_2\rangle \tag{1.15}$$

to mean that every amplitude in the complete set $|\psi_3\rangle$ is the sum of the corresponding amplitudes in the complete sets $|\psi_1\rangle$ and $|\psi_2\rangle$. This rule is exactly analogous to the rule for adding vectors because $\mathbf{b}_3 = \mathbf{b}_1 + \mathbf{b}_2$ implies that each component of $\mathbf{b}_3$ is the sum of the corresponding components of $\mathbf{b}_1$ and $\mathbf{b}_2$.

Since amplitudes are complex numbers, for any complex number α we can define

$$|\psi'\rangle = \alpha|\psi\rangle \tag{1.16}$$

to mean that every amplitude in the set $|\psi'\rangle$ is α times the corresponding amplitude in $|\psi\rangle$. Again there is an obvious parallel in the case of vectors: $3\mathbf{b}$ is the vector that has x component $3b_x$, etc.

1.3.3 Vector spaces and their adjoints

The analogy between kets and vectors proves extremely fruitful and is worth developing. For a mathematician, objects, like kets, that you can add and multiply by arbitrary complex numbers inhabit a **vector space**. Since we live in a (three-dimensional) vector space, we have a strong intuitive feel for the structures that arise in general vector spaces, and this intuition helps us to understand problems that arise with kets. Unfortunately our everyday experience does not prepare us for an important property of a general vector space, namely the existence of an associated 'adjoint' space, because the space adjoint to real three-dimensional space is indistinguishable from real space. In quantum mechanics and in relativity the two spaces are distinguishable. We now take a moment to develop the mathematical theory of general vector spaces in the context of kets in order to explain the relationship between a general vector space and its adjoint space. When we are merely using kets as examples of vectors, we shall call them "vectors". Appendix D explains how these ideas are relevant to relativity.

---

[1] Most elementary particles have intrinsic angular momentum or 'spin' (§7.4). A complete set of amplitudes for a particle such as electron or proton that has spin, includes information about the orientation of the spin. In the interests of simplicity, in our discussions particles are assumed to have no spin unless the contrary is explicitly stated, even though spinless particles are rather rare.

For any vector space V it is natural to choose a set of **basis vectors**, that is, a set of vectors $|i\rangle$ that is large enough for it to be possible to express any given vector $|\psi\rangle$ as a linear combination of the set's members. Specifically, for any ket $|\psi\rangle$ there are complex numbers a_i such that

$$|\psi\rangle = \sum_i a_i |i\rangle. \tag{1.17}$$

The set should be minimal in the sense that none of its members can be expressed as a linear combination of the remaining ones. In the case of ordinary three-dimensional space, basis vectors are provided by the unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ along the three coordinate axes, and any vector $\mathbf{b}$ can be expressed as the sum $\mathbf{b} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$, which is the analogue of equation (1.17).

In quantum mechanics an important role is played by complex-valued linear functions on the vector space V because these functions extract the amplitude for something to happen given that the system is in the state $|\psi\rangle$. Let $\langle f|$ (pronounced '**bra** f') be such a function. We denote by $\langle f|\psi\rangle$ the result of evaluating this function on the ket $|\psi\rangle$. Hence, $\langle f|\psi\rangle$ is a complex number (a probability amplitude) that in the ordinary notation of functions would be written $f\,(|\psi\rangle)$. The linearity of the function $\langle f|$ implies that for any complex numbers α, β and kets $|\psi\rangle$, $|\phi\rangle$, it is true that

$$\langle f|\big(\alpha|\psi\rangle + \beta|\phi\rangle\big) = \alpha\langle f|\psi\rangle + \beta\langle f|\phi\rangle. \tag{1.18}$$

Notice that the right side of this equation is a sum of two products of complex numbers, so it is well defined.

To define a function on V we have only to give a rule that enables us to evaluate the function on any vector in V. Hence we can define the sum $\langle h| \equiv \langle f| + \langle g|$ of two bras $\langle f|$ and $\langle g|$ by the rule

$$\langle h|\psi\rangle = \langle f|\psi\rangle + \langle g|\psi\rangle \tag{1.19}$$

Similarly, we define the bra $\langle p| \equiv \alpha\langle f|$ to be result of multiplying $\langle f|$ by some complex number α through the rule

$$\langle p|\psi\rangle = \alpha\langle f|\psi\rangle. \tag{1.20}$$

Since we now know what it means to add these functions and multiply them by complex numbers, they form a vector space V', called the **adjoint space** of V.

The **dimension** of a vector space is the number of vectors required to make up a basis for the space. We now show that V and V' have the same dimension. Let[2] $\{|i\rangle\}$ for $i = 1, N$ be a basis for V. Then a linear function $\langle f|$ on V is fully defined once we have given the N numbers $\langle f|i\rangle$. To see that this is true, we use (1.17) and the linearity of $\langle f|$ to calculate $\langle f|\psi\rangle$ for an arbitrary vector $|\psi\rangle = \sum_i a_i|i\rangle$:

$$\langle f|\psi\rangle = \sum_{i=1}^{N} a_i \langle f|i\rangle. \tag{1.21}$$

This result implies that we can define N functions $\langle j|$ $(j = 1, N)$ through the equations

$$\langle j|i\rangle = \delta_{ij}, \tag{1.22}$$

where δ_{ij} is 1 if $i = j$ and zero otherwise, because these equations specify the value that each bra $\langle j|$ takes on every basis vector $|i\rangle$ and therefore through

---

[2] Throughout this book the notation $\{x_i\}$ means 'the set of objects x_i'.

(1.21) the value that $\bar{j}$ takes on any vector ψ. Now consider the following linear combination of these bras:

$$\langle F| \equiv \sum_{j=1}^{N} \langle f|j\rangle\langle j|. \tag{1.23}$$

It is trivial to check that for any i we have $\langle F|i\rangle = \langle f|i\rangle$, and from this it follows that $\langle F| = \langle f|$ because we have already agreed that a bra is fully specified by the values it takes on the basis vectors. Since we have now shown that *any* bra can be expressed as a linear combination of the N bras specified by (1.22), and the latter are manifestly linearly independent, it follows that the dimensionality of V' is N, the dimensionality of V.

In summary, we have established that every N-dimensional vector space V comes with an N-dimensional space V' of linear functions on V, called the adjoint space. Moreover, we have shown that once we have chosen a basis $\{|i\rangle\}$ for V, there is an associated basis $\{\langle i|\}$ for V'. Equation (1.22) shows that there is an intimate relation between the ket $|i\rangle$ and the bra $\langle i|$: $\langle i|i\rangle = 1$ while $\langle j|i\rangle = 0$ for $j \neq i$. We acknowledge this relationship by saying that $\langle i|$ is the **adjoint** of $|i\rangle$. We extend this definition of an adjoint to an arbitrary ket $|\psi\rangle$ as follows: if

$$|\psi\rangle = \sum_i a_i|i\rangle \quad \text{then} \quad \langle\psi| \equiv \sum_i a_i^*\langle i|. \tag{1.24}$$

With this choice, when we evaluate the function $\langle\psi|$ on the ket $|\psi\rangle$ we find

$$\langle\psi|\psi\rangle = \left(\sum_i a_i^*\langle i|\right)\left(\sum_j a_j|j\rangle\right) = \sum_i |a_i|^2 \geq 0. \tag{1.25}$$

Thus for any state the number $\langle\psi|\psi\rangle$ is real and non-negative, and it can vanish only if $|\psi\rangle = 0$ because every a_i vanishes. We call this number the **length** of $|\psi\rangle$.

The components of an ordinary three-dimensional vector $\mathbf{b} = b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}$ are real. Consequently, we evaluate the length-square of $\mathbf{b}$ as simply $(b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}) \cdot (b_x\mathbf{i} + b_y\mathbf{j} + b_z\mathbf{k}) = b_x^2 + b_y^2 + b_z^2$. The vector on the extreme left of this expression is strictly speaking the adjoint of $\mathbf{b}$ but it is indistinguishable from it because we have not modified the components in any way. In the quantum mechanical case eq. 1.25, the components of the adjoint vector are complex conjugates of the components of the vector, so the difference between a vector and its adjoint is manifest.

If $|\phi\rangle = \sum_i b_i|i\rangle$ and $|\psi\rangle = \sum_i a_i|i\rangle$ are any two states, a calculation analogous to that in equation (1.25) shows that

$$\langle\phi|\psi\rangle = \sum_i b_i^* a_i, \tag{1.26}$$

where the b_i are the amplitudes that define the state $|\phi\rangle$. Similarly, we can show that $\langle\psi|\phi\rangle = \sum_i a_i^* b_i$, and from this it follows that

$$\langle\psi|\phi\rangle = \big(\langle\phi|\psi\rangle\big)^*. \tag{1.27}$$

We shall make frequent use of this equation.

Equation (1.26) shows that there is a close connection between extracting the complex number $\langle\phi|\psi\rangle$ from $\langle\phi|$ and $|\psi\rangle$ and the operation of taking the dot product between two vectors $\mathbf{b}$ and $\mathbf{a}$.

1.3.4 The energy representation

Suppose our system is a particle that is trapped in some potential well. Then the spectrum of allowed energies will be a set of discrete numbers E_0, $E_1, \ldots$ and a complete set of amplitudes are the amplitudes a_i whose mod squares give the probabilities p_i of measuring the energy to be E_i. Let $\{|i\rangle\}$ be a set of basis kets for the space V of the system's quantum states. Then we use the set of amplitudes a_i to associate them with a ket $|\psi\rangle$ through

$$|\psi\rangle = \sum_i a_i |i\rangle. \tag{1.28}$$

This equation relates a complete set of amplitudes $\{a_i\}$ to a certain ket $|\psi\rangle$. We discover the physical meaning of a particular basis ket, say $|k\rangle$, by examining the values that the expansion coefficients a_i take when we apply equation (1.28) in the case $|k\rangle = |\psi\rangle$. We clearly then have that $a_i = 0$ for $i \neq k$ and $a_k = 1$. Consequently, the quantum state $|k\rangle$ is that in which we are certain to measure the value E_k for the energy. We say that $|k\rangle$ is a state of **well defined** energy. It will help us remember this important identification if we relabel the basis kets, writing $|E_i\rangle$ instead of just $|i\rangle$, so that (1.28) becomes

$$|\psi\rangle = \sum_i a_i |E_i\rangle. \tag{1.29}$$

Suppose we multiply this equation through by $\langle E_k|$. Then by the linearity of this operation and the orthogonality relation (1.22) (which in our new notation reads $\langle E_k|E_i\rangle = \delta_{ik}$) we find

$$a_k = \langle E_k|\psi\rangle. \tag{1.30}$$

This is an enormously important result because it tells us how to extract from an arbitrary quantum state $|\psi\rangle$ the amplitude for finding that the energy is E_k.

Equation (1.25) yields

$$\langle\psi|\psi\rangle = \sum_i |a_i|^2 = \sum_i p_i = 1, \tag{1.31}$$

where the last equality follows because if we measure the energy, we must find *some* value, so the probabilities p_i must sum to unity. Thus kets that describe real quantum states must have unit length: we call kets with unit length **properly normalised**. During calculations we frequently encounter kets that are not properly normalised, and it is important to remember that the key rule (1.30) can be used to extract predictions only from properly normalised kets. Fortunately, any ket $|\phi\rangle = \sum_i b_i|i\rangle$ is readily normalised: it is straightforward to check that

$$|\psi\rangle \equiv \sum_i \frac{b_i}{\sqrt{\langle\phi|\phi\rangle}} |i\rangle \tag{1.32}$$

is properly normalised regardless of the values of the b_i.

1.3.5 Orientation of a spin-half particle

Formulae for the components of the spin angular momentum of a spin-half particle that we shall derive in §7.4.2 provide a nice illustration of how the abstract machinery just introduced enables us to predict the results of experiments.

If you measure one component, say s_z, of the spin $\mathbf{s}$ of an electron, you will obtain one of two results, either $s_z = \frac{1}{2}$ or $s_z = -\frac{1}{2}$. Moreover the state

$|+\rangle$ in which a measurement of s_z is certain to yield $\frac{1}{2}$ and the state $|-\rangle$ in which the measurement is certain to yield $-\frac{1}{2}$ form a complete set of states for the electron's spin. That is, *any* state of spin can be expressed as a linear combination of $|+\rangle$ and $|-\rangle$:

$$|\psi\rangle = a_-|-\rangle + a_+|+\rangle. \tag{1.33}$$

Let $\mathbf{n}$ be the unit vector in the direction with polar coordinates (θ, ϕ). Then the state $|+, \mathbf{n}\rangle$ in which a measurement of the component of $\mathbf{s}$ along $\mathbf{n}$ is certain to return $\frac{1}{2}$ turns out to be (Problem 7.12)

$$|+, \mathbf{n}\rangle = \sin(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi/2}|-\rangle + \cos(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi/2}|+\rangle. \tag{1.34a}$$

Similarly the state $|-, \mathbf{n}\rangle$ in which a measurement of the component of $\mathbf{s}$ along $\mathbf{n}$ is certain to return $-\frac{1}{2}$ is

$$|-, \mathbf{n}\rangle = \cos(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi/2}|-\rangle - \sin(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi/2}|+\rangle. \tag{1.34b}$$

By equation (1.24) the adjoints of these kets are the bras

$$\begin{aligned}\langle +, \mathbf{n}| &= \sin(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi/2}\langle -| + \cos(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi/2}\langle +|\\ \langle -, \mathbf{n}| &= \cos(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi/2}\langle -| - \sin(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi/2}\langle +|.\end{aligned} \tag{1.35}$$

From these expressions it is easy to check that the kets $|\pm, \mathbf{n}\rangle$ are properly normalised and orthogonal to one another.

Suppose we have just measured s_z and found the value to be $\frac{1}{2}$ and we want the amplitude $A_-(\mathbf{n})$ to find $-\frac{1}{2}$ when we measure $\mathbf{n}\cdot\mathbf{s}$. Then the state of the system is $|\psi\rangle = |+\rangle$ and the required amplitude is

$$A_-(\mathbf{n}) = \langle -, \mathbf{n}|\psi\rangle = \langle -, \mathbf{n}|+\rangle = -\sin(\theta/2)\mathrm{e}^{\mathrm{i}\phi/2}, \tag{1.36}$$

so the probability of this outcome is

$$P_-(\mathbf{n}) = |A_-(\mathbf{n})|^2 = \sin^2(\theta/2). \tag{1.37}$$

This vanishes when $\theta = 0$ as it should since then $\mathbf{n} = (0, 0, 1)$ so $\mathbf{n}\cdot\mathbf{s} = s_z$, and we are guaranteed to find $s_z = \frac{1}{2}$ rather than $-\frac{1}{2}$. $P_-(\mathbf{n})$ rises to $\frac{1}{2}$ when $\theta = \pi/2$ and $\mathbf{n}$ lies somewhere in the x, y plane. In particular, if $s_z = \frac{1}{2}$, a measurement of s_x is equally likely to return either of the two possible values $\pm\frac{1}{2}$.

Putting $\theta = \pi/2$, $\phi = 0$ into equations (1.34) we obtain expressions for the states in which the result of a measurement of s_x is certain

$$|+, x\rangle = \frac{1}{\sqrt{2}}\left(|-\rangle + |+\rangle\right) \quad ; \quad |-, x\rangle = \frac{1}{\sqrt{2}}\left(|-\rangle - |+\rangle\right). \tag{1.38}$$

Similarly, inserting $\theta = \pi/2$, $\phi = \pi/2$ we obtain the states in which the result of measuring s_y is certain

$$|+, y\rangle = \frac{\mathrm{e}^{\mathrm{i}\pi/4}}{\sqrt{2}}\left(|-\rangle - \mathrm{i}|+\rangle\right) \quad ; \quad |-, y\rangle = \frac{\mathrm{e}^{\mathrm{i}\pi/4}}{\sqrt{2}}\left(|-\rangle + \mathrm{i}|+\rangle\right). \tag{1.39}$$

Notice that $|+, x\rangle$ and $|+, y\rangle$ are both states in which the probability of measuring s_z to be $\frac{1}{2}$ is $\frac{1}{2}$. What makes them physically distinct states is that the ratio of the *amplitudes* to measure $\pm\frac{1}{2}$ for s_z is unity in one case and i in the other.

1.3.6 Polarisation of photons

A discussion of the possible polarisations of a beam of light displays an interesting connection between quantum amplitudes and classical physics. At any instant in a polarised beam of light, the electric vector $\mathbf{E}$ is in one particular direction perpendicular to the beam. In a plane-polarised beam, the direction of $\mathbf{E}$ stays the same, while in a circularly polarised beam it rotates. A sheet of Polaroid transmits the component of $\mathbf{E}$ in one direction and blocks the perpendicular component. Consequently, in the transmitted beam $|\mathbf{E}|$ is smaller than in the incident beam by a factor $\cos\theta$, where θ is the angle between the incident field and the direction in the Polaroid that transmits the field. Since the beam's energy flux is proportional to $|\mathbf{E}|^2$, a fraction $\cos^2\theta$ of the beam's energy is transmitted by the Polaroid.

Individual photons either pass through the Polaroid intact or are absorbed by it depending on which quantum state they are found to be in when they are 'measured' by the Polaroid. Let $|{\rightarrow}\rangle$ be the state in which the photon will be transmitted and $|{\uparrow}\rangle$ that in which it will be blocked. Then the photons of the incoming plane-polarised beam are in the state

$$|\psi\rangle = \cos\theta|{\rightarrow}\rangle + \sin\theta|{\uparrow}\rangle, \tag{1.40}$$

so each photon has an amplitude $a_{\rightarrow} = \cos\theta$ for a measurement by the Polaroid to find it in the state $|{\rightarrow}\rangle$ and be transmitted, and an amplitude $a_{\uparrow} = \sin\theta$ to be found to be in the state $|{\uparrow}\rangle$ and be blocked. The fraction of the beam's photons that are transmitted is the probability get through $P_{\rightarrow} = |a_{\rightarrow}|^2 = \cos^2\theta$. Consequently a fraction $\cos^2\theta$ of the incident energy is transmitted, in agreement with classical physics.

The states $|{\rightarrow}\rangle$ and $|{\uparrow}\rangle$ form a complete set of states for photons that move in the direction of the beam. An alternative complete set of states is the set $\{|+\rangle, |-\rangle\}$ formed by the state $|+\rangle$ of a right-hand circularly polarised photon and the state $|-\rangle$ of a left-hand circularly polarised photon. In the laboratory a circularly polarised beam is often formed by passing a plane polarised beam through a birefringent material such as calcite that has its axes aligned at 45° to the incoming plane of polarisation. The incoming beam is resolved into its components parallel to the calcite's axes, and one component is shifted in phase by $\pi/2$ with respect to the other. In terms of unit vectors $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$ parallel to the calcite's axes, the incoming field is

$$\mathbf{E} = \frac{E}{\sqrt{2}}\Re\left\{(\hat{\mathbf{e}}_x + \hat{\mathbf{e}}_y)\mathrm{e}^{-\mathrm{i}\omega t}\right\} \tag{1.41}$$

and the outgoing field of a left-hand polarised beam is

$$\mathbf{E}_- = \frac{E}{\sqrt{2}}\Re\left\{(\hat{\mathbf{e}}_x + \mathrm{i}\hat{\mathbf{e}}_y)\mathrm{e}^{-\mathrm{i}\omega t}\right\}, \tag{1.42a}$$

while the field of a right-hand polarised beam would be

$$\mathbf{E}_+ = \frac{E}{\sqrt{2}}\Re\left\{(\hat{\mathbf{e}}_x - \mathrm{i}\hat{\mathbf{e}}_y)\mathrm{e}^{-\mathrm{i}\omega t}\right\}. \tag{1.42b}$$

The last two equations express the electric field of a circularly polarised beam as a linear combination of plane polarised beams that differ in phase. Conversely, by adding (1.42b) to equation (1.42a), we can express the electric field of a beam polarised along the x axis as a linear combination of the fields of two circularly-polarised beams.

Similarly, the quantum state of a circularly polarised photon is a linear superposition of linearly-polarised quantum states:

$$|\pm\rangle = \frac{1}{\sqrt{2}}\left(|{\rightarrow}\rangle \mp \mathrm{i}|{\uparrow}\rangle\right), \tag{1.43}$$

and conversely, a state of linear polarisation is a linear superposition of states of circular polarisation:

$$|\!\rightarrow\rangle = \frac{1}{\sqrt{2}}\left(|+\rangle + |-\rangle\right). \tag{1.44}$$

Whereas in classical physics complex numbers are just a convenient way of representing the real function $\cos(\omega t + \phi)$ for arbitrary phase ϕ, quantum amplitudes are inherently complex and the operator $\Re$ is not used. Whereas in classical physics a beam may be linearly polarised in a particular direction, or circularly polarised in a given sense, in quantum mechanics an individual photon has an amplitude to be linearly polarised in a any chosen direction *and* an amplitude to be circularly polarised in a given sense. The amplitude to be linearly polarised may vanish in one particular direction, *or* it may vanish for one sense of circular polarisation. In the general case the photon will have a non-vanishing amplitude to be polarised in any direction and any sense. After it has been transmitted by an analyser such as Polaroid, it will certainly be in whatever state the analyser transmits.

1.4 Measurement

Equation (1.28) expresses the quantum state of a system $|\psi\rangle$ as a sum over states in which a particular measurement, such as energy, is certain to yield a specified value. The coefficients in this expansion yield as their mod-squares the probabilities with which the possible results of the measurement will be obtained. Hence so long as there is more than one term in the sum, the result of the measurement is in doubt. This uncertainty does not reflect shortcomings in the measuring apparatus, but is inherent in the physical situation – any defects in the measuring apparatus will increase the uncertainty above the irreducible minimum implied by the expansion coefficients, and in §6.3 the theory will be adapted to include such additional uncertainty.

Here we are dealing with ideal measurements, and such measurements are reproducible. Therefore, if a second measurement is made immediately after the first, the same result will be obtained. From this observation it follows that the quantum state of the system is changed by the first measurement from $|\psi\rangle = \sum_i a_i|i\rangle$ to $|\psi\rangle = |I\rangle$, where $|I\rangle$ is the state in which the measurement is guaranteed to yield the value that was obtained by the first measurement. The abrupt change in the quantum state from $\sum_i a_i|i\rangle$ to $|I\rangle$ that accompanies a measurement is referred to as the **collapse of the wavefunction**.

What happens when the "wavefunction collapses"? It is tempting to suppose that this event is not a physical one but merely an updating of our knowledge of the system: that the system was already in the state $|I\rangle$ before the measurement, but we only became aware of this fact when the measurement was made. It turns out that this interpretation is untenable, and that wavefunction collapse is associated with a real physical disturbance of the system. This topic is explored further in §6.5.

Problems

1.1 What physical phenomenon requires us to work with probability amplitudes rather than just with probabilities, as in other fields of endeavour?

1.2 What properties cause complete sets of amplitudes to constitute the elements of a vector space?

1.3 V' is the adjoint space of the vector space V. For a mathematician, what objects comprise V'?

1.4 In quantum mechanics, what objects are the members of the vector space V? Give an example for the case of quantum mechanics of a member of the adjoint space V' and explain how members of V' enable us to predict the outcomes of experiments.

1.5 Given that $|\psi\rangle = \mathrm{e}^{\mathrm{i}\pi/5}|a\rangle + \mathrm{e}^{\mathrm{i}\pi/4}|b\rangle$, express $\langle\psi|$ as a linear combination of $\langle a|$ and $\langle b|$.

1.6 What properties characterise the bra $\langle a|$ that is associated with the ket $|a\rangle$?

1.7 An electron can be in one of two potential wells that are so close that it can "tunnel" from one to the other. Its state vector can be written

$$|\psi\rangle = a|A\rangle + b|B\rangle, \tag{1.45}$$

where $|A\rangle$ is the state of being in the first well and $|B\rangle$ is the state of being in the second well and all kets are correctly normalised. What is the probability of finding the particle in the first well given that: (a) $a = \mathrm{i}/2$; (b) $b = \mathrm{e}^{\mathrm{i}\pi}$; (c) $b = \frac{1}{3} + \mathrm{i}/\sqrt{2}$?

1.8 An electron can "tunnel" between potential wells that form a chain, so its state vector can be written

$$|\psi\rangle = \sum_{-\infty}^{\infty} a_n|n\rangle, \tag{1.46a}$$

where $|n\rangle$ is the state of being in the n^{th} well, where n increases from left to right. Let

$$a_n = \frac{1}{\sqrt{2}}\left(\frac{-\mathrm{i}}{3}\right)^{|n|/2} \mathrm{e}^{\mathrm{i}n\pi}. \tag{1.46b}$$

a. What is the probability of finding the electron in the n^{th} well?
b. What is the probability of finding the electron in well 0 or anywhere to the right of it?

2

Operators, measurement and time evolution

In the last chapter we saw that each quantum state of a system is represented by a point or 'ket' $|\psi\rangle$ that lies in an abstract vector space. We saw that states for which there is no uncertainty in the value that will be measured for a quantity such as energy, form a set of basis states for this space – these basis states are analogous to the unit vectors **i**, **j** and **k** of ordinary vector geometry. In this chapter we develop these ideas further by showing how every measurable quantity such as position, momentum or energy is associated with an operator on state space. We shall see that the energy operator plays a special role in that it determines how a system's ket $|\psi\rangle$ moves through state space over time. Using these operators we are able at the end of the chapter to study the dynamics of a free particle, and to understand how the uncertainties in the position and momentum of a particle are intimately connected with one another, and how they evolve in time.

2.1 Operators

A **linear operator** on the vector space V is an object Q that transforms kets into kets in a linear way. That is, if $|\psi\rangle$ is a ket, then $|\phi\rangle = Q|\psi\rangle$ is another ket, and if $|\chi\rangle$ is a third ket and α and β are complex numbers, we have

$$Q\big(\alpha|\psi\rangle + \beta|\chi\rangle\big) = \alpha(Q|\psi\rangle) + \beta(Q|\chi\rangle). \tag{2.1}$$

Consider now the linear operator

$$I = \sum_i |i\rangle\langle i|, \tag{2.2}$$

where $\{|i\rangle\}$ is any set of basis kets. I really *is* an operator because if we apply it to any ket $|\psi\rangle$, we get a linear combination of kets, which must itself be a ket:

$$I|\psi\rangle = \sum_i |i\rangle\langle i|\psi\rangle = \sum_i (\langle i|\psi\rangle)\,|i\rangle, \tag{2.3}$$

where we are able to move $\langle i|\psi\rangle$ around freely because it's just a complex number. To determine which ket $I|\psi\rangle$ is, we substitute into (2.3) the expansion (1.17) of $|\psi\rangle$ and use the orthogonality relation (1.22):

$$\begin{aligned} I|\psi\rangle &= \sum_i |i\rangle\langle i|\Big(\sum_j a_j|j\rangle\Big) \\ &= \sum_i a_i|i\rangle = |\psi\rangle. \end{aligned} \tag{2.4}$$

We have shown that I applied to an arbitrary ket $|\psi\rangle$ yields that same ket. Hence I is the **identity operator**. We shall make extensive use of this fact.

Consider now the operator

$$H = \sum_i E_i|E_i\rangle\langle E_i|. \tag{2.5}$$

This is the most important single operator in quantum mechanics. It is called the **Hamiltonian** in honour of W.R. Hamilton, who introduced its classical analogue.[1] We use H to operate on an arbitrary ket $|\psi\rangle$ to form the ket $H|\psi\rangle$, and then we bra through by the adjoint $\langle\psi|$ of $|\psi\rangle$. We have

$$\langle\psi|H|\psi\rangle = \sum_i E_i\langle\psi|E_i\rangle\langle E_i|\psi\rangle. \tag{2.6}$$

By equation (1.29) $\langle E_i|\psi\rangle = a_i$, while by (1.24) $\langle\psi|E_i\rangle = a_i^*$. Thus

$$\langle\psi|H|\psi\rangle = \sum_i E_i|a_i|^2 = \sum_i p_iE_i = \langle E\rangle\,. \tag{2.7}$$

Here is yet another result of fundamental importance: if we squeeze the Hamiltonian between a quantum state $|\psi\rangle$ and its adjoint bra, we obtain the expectation value of the energy for that state.

It is straightforward to generalise this result for the expectation value of the energy to other measurable quantities: if Q is something that we can measure (often called an **observable**) and its spectrum of possible values is $\{q_i\}$, then we expand an arbitrary ket $|\psi\rangle$ as a linear combination of states $|q_i\rangle$ in which the value of Q is well defined,

$$|\psi\rangle = \sum_i a_i|q_i\rangle, \tag{2.8}$$

and with Q we associate the operator

$$Q = \sum_i q_i|q_i\rangle\langle q_i|. \tag{2.9}$$

Then $\langle\psi|Q|\psi\rangle$ is the expectation value of Q when our system is in the state $|\psi\rangle$. When the state in question is obvious from the context, we shall sometimes write the expectation value of Q simply as $\langle Q\rangle$.

When a linear operator R turns up in any mathematical problem, it is generally expedient to investigate its eigenvalues and eigenvectors. An **eigenvector** is a vector that R simply rescales, and its **eigenvalue** is the rescaling factor. Thus, let $|r\rangle$ be an eigenvector of R, and r be its eigenvalue, then we have

$$R|r\rangle = r|r\rangle. \tag{2.10}$$

---

[1] William Rowan Hamilton (1805–1865) was a protestant Irishman who was appointed the Andrews' Professor of Astronomy at Trinity College Dublin while still an undergraduate. Although he did not contribute to astronomy, he made important contributions to optics and mechanics, and to pure mathematics with his invention of quaternions, the first non-commutative algebra.

Box 2.1: Hermitian Operators

Let Q be a Hermitian operator with eigenvalues q_i and eigenvectors $|q_i\rangle$. Then we bra the defining equation of $|q_i\rangle$ through by $\langle q_k|$, and bra the defining equation of $|q_k\rangle$ through by $\langle q_i|$:

$$\langle q_k|Q|q_i\rangle = q_i\langle q_k|q_i\rangle \qquad \langle q_i|Q|q_k\rangle = q_k\langle q_i|q_k\rangle.$$

We next take the complex conjugate of the second equation from the first. The left side then vanishes because Q is Hermitian, so with equation (1.27)

$$0 = (q_i - q_k^*)\langle q_k|q_i\rangle.$$

Setting $k = i$ we find that $q_i = q_i^*$ since $\langle q_i|q_i\rangle > 0$. Hence the eigenvalues are real. When $q_i \neq q_k$, we must have $\langle q_k|q_i\rangle = 0$, so the eigenvectors belonging to distinct eigenvalues are orthogonal.

What are the eigenvectors and eigenvalues of H? If we apply H to $|E_k\rangle$, we find

$$H|E_k\rangle = \sum_i E_i|E_i\rangle\langle E_i|E_k\rangle = E_k|E_k\rangle. \tag{2.11}$$

So the eigenvectors of H are the states of well defined energy, and its eigenvalues are the possible results of a measurement of energy. Clearly this important result generalises immediately to eigenvectors and eigenvalues of the operator Q that we have associated with an arbitrary observable.

Consider the complex number $\langle\phi|Q|\psi\rangle$, where $|\phi\rangle$ and $|\psi\rangle$ are two arbitrary quantum states. After expanding the states in terms of the eigenvectors of Q, we have

$$\langle\phi|Q|\psi\rangle = \Big(\sum_i b_i^*\langle q_i|\Big)Q\Big(\sum_j a_j|q_j\rangle\Big) = \sum_{ij} b_i^* a_j q_j \delta_{ij} = \sum_i b_i^* q_i a_i \tag{2.12}$$

Similarly, $\langle\psi|Q|\phi\rangle = \sum_i a_i^* q_i b_i$. Hence so long as the spectrum $\{q_i\}$ of Q consists entirely of real numbers (which is physically reasonable), then

$$(\langle\phi|Q|\psi\rangle)^* = \langle\psi|Q|\phi\rangle \tag{2.13}$$

for any two states $|\phi\rangle$ and $|\psi\rangle$. An operator with this property is said to be **Hermitian**. Hermitian operators have nice properties. In particular, one can prove – see Box 2.1 – that they have real eigenvalues and mutually orthogonal eigenvectors, and it is because we require these properties on physical grounds that the operators of observables turn out to be Hermitian. In Chapter 4 we shall find that Hermitian operators arise naturally from another physical point of view.

Although the operators associated with observables are always Hermitian, operators that are not Hermitian turn out to be extremely useful. With a non-Hermitian operator R we associate another operator $R^\dagger$ called its **Hermitian adjoint** by requiring that for any states $|\phi\rangle$ and $|\psi\rangle$ it is true that

$$\big(\langle\phi|R^\dagger|\psi\rangle\big)^* = \langle\psi|R|\phi\rangle. \tag{2.14}$$

Comparing this equation with equation (2.13) it is clear that a Hermitian operator Q is its own adjoint: $Q^\dagger = Q$.

By expanding the kets $|\phi\rangle$ and $|\psi\rangle$ in the equation $|\phi\rangle = R|\psi\rangle$ as sums of basis kets, we show that R is completely determined by the array of numbers (called **matrix elements**)

$$R_{ij} \equiv \langle i|R|j\rangle. \tag{2.15}$$

Table 2.1 Rules for Hermitian adjoints

Object	i	$\vert\psi\rangle$	R	QR	$R\vert\psi\rangle$	$\langle\phi\vert R\vert\psi\rangle$
Adjoint	−i	$\langle\psi\vert$	$R^\dagger$	$R^\dagger Q^\dagger$	$\langle\psi\vert R^\dagger$	$\langle\psi\vert R^\dagger\vert\phi\rangle$

In fact

$$\begin{aligned}|\phi\rangle = \sum_i b_i|i\rangle = R|\psi\rangle = \sum_j a_j R|j\rangle \\ \Rightarrow \quad b_i = \sum_j a_j\langle i|R|j\rangle = \sum_j R_{ij}a_j.\end{aligned} \tag{2.16}$$

If in equation (2.14) we set $|\phi\rangle = |i\rangle$ and $|\psi\rangle = |j\rangle$, we discover the relation between the matrix of R and that of $R^\dagger$:

$$(R^\dagger_{ij})^* = R_{ji} \quad \Leftrightarrow \quad R^\dagger_{ij} = R^*_{ji}. \tag{2.17}$$

Hence the matrix of $R^\dagger$ is the complex-conjugate transpose of the matrix for R. If R is Hermitian so that $R^\dagger = R$, the matrix R_{ij} must equal its complex-conjugate transpose, that is, it must be an Hermitian matrix.

Operators can be multiplied together: when the operator QR operates on $|\psi\rangle$, the result is what you get by operating first with R and then applying Q to $R|\psi\rangle$. We shall frequently need to find the Hermitian adjoints of such products. To find out how to do this we replace R in (2.17) by QR:

$$(QR)^\dagger_{ij} = (QR)^*_{ji} = \sum_k Q^*_{jk}R^*_{ki} = \sum_k R^\dagger_{ik}Q^\dagger_{kj} = (R^\dagger Q^\dagger)_{ij}. \tag{2.18}$$

Thus, to dagger a product we reverse the terms and dagger the individual operators. By induction it is now easy to show that

$$(ABC\ldots Z)^\dagger = Z^\dagger \ldots C^\dagger B^\dagger A^\dagger. \tag{2.19}$$

If we agree that the Hermitian adjoint of a complex number is its complex conjugate and that $|\psi\rangle^\dagger \equiv \langle\psi|$ and $\langle\psi|^\dagger \equiv |\psi\rangle$, then we can consider the basic rule (2.14) for taking the complex conjugate of a matrix element to be a generalisation of the rule we have derived about reversing the order and daggering the components of a product of operators. The rules for taking Hermitian adjoints are summarised in Table 2.1.

Functions of operators We shall frequently need to evaluate functions of operators. For example, the potential energy of a particle is a function $V(\hat{x})$ of the position operator $\hat{x}$. Let f be any function of one variable and R be any operator. Then we define the operator $f(R)$ by the equation

$$f(R) \equiv \sum_i f(r_i)|r_i\rangle\langle r_i|, \tag{2.20}$$

where the r_i and $|r_i\rangle$ are the eigenvalues and eigenkets of R. This definition defines $f(R)$ to be the operator that has the same eigenkets as R and the eigenvalues that you get by evaluating the function f on the eigenvalues of R.

Commutators The commutator of two operators A, B is defined to be

$$[A, B] \equiv AB - BA. \tag{2.21}$$

If $[A, B] \neq 0$, it is impossible to find a complete set of mutual eigenkets of A and B (Problem 2.20). Conversely, it can be shown that if $[A, B] = 0$ there *is* a complete set of mutual eigenkets of A and B, that is, there is a complete set of states of the system in which there is no uncertainty in the value that will be obtained for either A or B. We shall make extensive use of this fact.

Notice that the word *complete* appears in both these statements; even in the case $[A,B] \neq 0$ it may be possible to find states in which both A and B have definite values. It is just that such states cannot form a complete set. Similarly, when $[A,B] = 0$ there can be states for which A has a definite value but B does not. The literature is full of inaccurate statements about the implications of $[A,B]$ being zero or non-zero.

Three invaluable rules are

$$\begin{aligned} [A+B,C] &= [A,C]+[B,C] \\ AB &= BA+[A,B] \\ [AB,C] &= [A,C]B+A[B,C]. \end{aligned} \tag{2.22}$$

All three rules are trivial to prove by explicitly writing out the contents of the square brackets. With these rules it is rarely necessary to write out the contents of a commutator again, so they eliminate a common source of error and tedium in calculations. Notice the similarity of the third rule to the standard rule for differentiating a product: $\mathrm{d}(ab)/\mathrm{d}c = (\mathrm{d}a/\mathrm{d}c)b + a(\mathrm{d}b/\mathrm{d}c)$. The rule is easily generalised by induction to the rule

$$[ABC\ldots,Z] = [A,Z]BC\ldots + A[B,Z]C\ldots + \ldots + AB[C,Z]\ldots \tag{2.23}$$

We shall frequently need to evaluate the commutator of an operator A with a function f of an operator B. We assume that f has a convergent Taylor series[2] $f = f_0 + f'B + \frac{1}{2}f''B^2 + \cdots$, where $f_0 \equiv f(0)$, $f' \equiv (\mathrm{d}f(x)/\mathrm{d}x)_0$, etc., are numbers. Then

$$\begin{aligned} [A,f(B)] = f'[A,B] &+ \tfrac{1}{2}f''([A,B]B + B[A,B]) \\ &+ \tfrac{1}{3!}f'''([A,B]B^2 + B[A,B]B + B^2[A,B]) + \cdots \end{aligned} \tag{2.24}$$

In the important case in which B commutes with $[A,B]$, this expression simplifies dramatically

$$[A,f(B)] = [A,B](f' + f''B + \tfrac{1}{2}f'''B^2 + \cdots) = [A,B]\frac{\mathrm{d}f}{\mathrm{d}B}. \tag{2.25}$$

We shall use this formula several times.

2.2 Evolution in time

Since physics is about predicting the future, equations of motion lie at its heart. Newtonian dynamics is dominated by the equation of motion $\mathbf{f} = m\mathbf{a}$, where $\mathbf{f}$ is the force on a particle of mass m and $\mathbf{a}$ is the resulting acceleration. In quantum mechanics the analogous dynamical equation is the **time-dependent Schrödinger equation (TDSE)**:[3]

$$\mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial t} = H|\psi\rangle. \tag{2.26}$$

For future reference we use the rules of Table 2.1 to derive from this equation the equation of motion of a bra:

$$-\mathrm{i}\hbar\frac{\partial\langle\psi|}{\partial t} = \langle\psi|H, \tag{2.27}$$

---

[2] If necessary, we expand $f(x)$ about some point $x_0 \neq 0$, i.e., in powers of $x - x_0$, so we don't need to worry that the series about the origin may not converge for all x.

[3] Beginners sometimes interpret the TDSE as stating that $H = \mathrm{i}\hbar\partial/\partial t$. This is as unhelpful as interpreting $\mathbf{f} = m\mathbf{a}$ as a definition of $\mathbf{f}$. For Newton's equation to be useful it has to be supplemented by a description of the forces acting on the particle. Similarly, the TDSE is useful only when we have another expression for H.

where we have used the fact that H is Hermitian, so $H^\dagger = H$. The great importance of the Hamiltonian operator is due to its appearance in the TDSE, which must be satisfied by the ket of *any* system. We shall see below in several concrete examples that the TDSE, which we have not attempted to motivate physically, generates familiar motions in circumstances that permit classical mechanics to be used.

One perhaps surprising aspect of the TDSE we can justify straight away: while Newton's second law is a second-order differential equation, the TDSE is first-order. Since it is first order, the boundary data at $t = 0$ required to solve for $|\psi, t\rangle$ at $t > 0$ comprise the ket $|\psi, 0\rangle$. If the equation were second-order in time, like Newton's law, the required boundary data would include $\partial|\psi\rangle/\partial t$. But $|\psi, 0\rangle$ by hypothesis constitutes a complete set of amplitudes; it embodies *everything* we know about the current state of the system. If mathematics required us to know something about the system in addition to $|\psi, 0\rangle$, then either $|\psi\rangle$ would not constitute a complete set of amplitudes, or physics could offer no hope of predicting the future, and it would be time to take up biology or accountancy, or whatever.

The TDSE tells us that states of well-defined energy evolve in time in an exceptionally simple way

$$\mathrm{i}\hbar\frac{\partial|E_n\rangle}{\partial t} = H|E_n\rangle = E_n|E_n\rangle, \tag{2.28}$$

which implies that

$$|E_n, t\rangle = |E_n, 0\rangle\mathrm{e}^{-\mathrm{i}E_n t/\hbar}. \tag{2.29}$$

That is, the passage of time simply changes the phase of the ket at a rate $E_n/\hbar$.

We can use this result to calculate the time evolution of an arbitrary state $|\psi\rangle$. In the energy representation the state is

$$|\psi, t\rangle = \sum_n a_n(t)|E_n, t\rangle. \tag{2.30}$$

Substituting this expansion into the TDSE (2.26) we find

$$\mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial t} = \sum_n \mathrm{i}\hbar\left(\dot{a}_n|E_n\rangle + a_n\frac{\partial|E_n\rangle}{\partial t}\right) = \sum_n a_n H|E_n\rangle, \tag{2.31}$$

where a dot denotes differentiation with respect to time. The right side cancels with the second term in the middle, so we have $\dot{a}_n = 0$. Since the a_n are constant, on eliminating $|E_n, t\rangle$ between equations (2.29) and (2.30), we find that the evolution of $|\psi\rangle$ is simply given by

$$|\psi, t\rangle = \sum_n a_n \mathrm{e}^{-\mathrm{i}E_n t/\hbar}|E_n, 0\rangle. \tag{2.32}$$

We shall use this result time and again.

States of well-defined energy are unphysical and never occur in Nature because they are incapable of changing in any way, and hence it is impossible to get a system into such a state. But they play an extremely important role in quantum mechanics because they provide the almost trivial solution (2.32) to the governing equation of the theory, (2.26). Given the central role of these states, we spend much time solving their defining equation

$$H|E_n\rangle = E_n|E_n\rangle, \tag{2.33}$$

which is known as the **time-independent Schrödinger equation**, or **TISE** for short.

2.2.1 Evolution of expectation values

We have seen that $\langle\psi|Q|\psi\rangle$ is the expectation value of the observable Q when the system is in the state $|\psi\rangle$, and that expectation values provide a natural connection to classical physics, which is about situations in which the result of a measurement is almost certain to lie very close to the quantum-mechanical expectation value. We can use the TDSE to determine the rate of change of this expectation value:

$$\begin{aligned} \mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\langle\psi|Q|\psi\rangle &= -\langle\psi|HQ|\psi\rangle + \mathrm{i}\hbar\langle\psi|\frac{\partial Q}{\partial t}|\psi\rangle + \langle\psi|QH|\psi\rangle \\ &= \langle\psi|[Q,H]|\psi\rangle + \mathrm{i}\hbar\langle\psi|\frac{\partial Q}{\partial t}|\psi\rangle, \end{aligned} \tag{2.34}$$

where we have used both the TDSE (2.26) and its Hermitian adjoint (2.27) and the square bracket denotes a commutator – see (2.21). Usually operators are independent of time (i.e., $\partial Q/\partial t = 0$), and then the rate of change of an expectation value is the expectation value of the operator $-\mathrm{i}[Q,H]/\hbar$. This important result is known as **Ehrenfest's theorem**.

If a time-independent operator Q happens to commute with the Hamiltonian, that is if $[Q,H] = 0$, then for any state $|\psi\rangle$ the expectation value of Q is constant in time, or a **conserved quantity**. Moreover, in these circumstances Q^2 also commutes with H, so $\langle\psi|(\Delta Q)^2|\psi\rangle = \left\langle Q^2\right\rangle - \left\langle Q\right\rangle^2$ is also constant. If initially ψ is a state of well-defined Q, i.e., $|\psi\rangle = |q_i\rangle$ for some i, then $\left\langle(\Delta Q)^2\right\rangle = 0$ at all times. Hence, whenever $[Q,H] = 0$, a state of well defined Q evolves into another such state, so the value of Q can be known precisely at all times. The value q_i is then said to be a **good quantum number**. We always need to label states in some way. The label should be something that can be checked at any time and is not constantly changing. Good quantum numbers have precisely these properties, so they are much employed as labels of states.

If the system is in a state of well defined energy, the expectation value of *any* time-independent operator is time-independent, even if the operator does not commute with H. This is true because in these circumstances equation (2.34) becomes

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\langle E|Q|E\rangle = \langle E|(QH - HQ)|E\rangle = (E - E)\langle E|Q|E\rangle = 0, \tag{2.35}$$

where we have used the equation $H|E\rangle = E|E\rangle$ and its Hermitian adjoint. In view of this property of having constant expectation values of all time-independent operators, states of well defined energy are called **stationary states**.

Since H inevitably commutes with itself, equation (2.34) gives for the rate of change of the expectation of the energy

$$\frac{\mathrm{d}\left\langle E\right\rangle}{\mathrm{d}t} = \left\langle\frac{\partial H}{\partial t}\right\rangle. \tag{2.36}$$

In particular $\langle E\rangle$ is constant if the Hamiltonian is time-independent. This is a statement of the principle of the conservation of energy since time-dependence of the Hamiltonian arises only when some external force is working on the system. For example, a particle that is gyrating in a time-dependent magnetic field has a time-dependent Hamiltonian because work is being done either on or by the currents that generate the field.

2.3 The position representation

If the system consists of a single particle that can move in only one dimension, the amplitudes $\psi(x)$ to find the particle at x for x in $(-\infty, \infty)$ constitute a complete set of amplitudes. By analogy with equation (1.29) we have[4]

$$|\psi\rangle = \int_{-\infty}^{\infty} \mathrm{d}x\, \psi(x)|x\rangle. \tag{2.37}$$

Here an integral replaces the sum because the spectrum of possible values of x is continuous rather than discrete. Our basis kets are the states $|x\rangle$ in which the particle is definitely at x. By analogy with equation (1.30) we have

$$\psi(x) = \langle x|\psi\rangle. \tag{2.38}$$

Notice that both sides of this equation are complex numbers that depend on the variable x, that is, they are complex-valued functions of x. For historical reasons, the function $\psi(x)$ is called the **wavefunction** of the particle. By the usual rule (1.27) for complex conjugation of a bra-ket we have

$$\psi^*(x) = \langle\psi|x\rangle. \tag{2.39}$$

The analogue for the kets $|x\rangle$ of the orthogonality relation (1.22) is

$$\langle x'|x\rangle = \delta(x - x'), \tag{2.40}$$

where the **Dirac delta function** $\delta(x - x')$ is zero for $x \neq x'$ because when the particle is at x, it has zero amplitude to be at a different location x'. We get insight into the value of $\delta(x - x')$ for $x = x'$ by multiplying equation (2.37) through by $\langle x'|$ and using equation (2.38) to eliminate $\langle x'|\psi\rangle$:

$$\begin{aligned}\langle x'|\psi\rangle = \psi(x') &= \int \mathrm{d}x\, \psi(x)\langle x'|x\rangle \\ &= \int \mathrm{d}x\, \psi(x)\delta(x - x').\end{aligned} \tag{2.41}$$

Since $\delta(x - x')$ is zero for $x \neq x'$, we can replace $\psi(x)$ in the integrand by $\psi(x')$ and then take this number outside the integral sign and cancel it with the $\psi(x')$ on the left hand side. What remains is the equation

$$1 = \int \mathrm{d}x\, \delta(x - x'). \tag{2.42}$$

Thus there is unit area under the graph of $\delta(x)$, which is remarkable, given that the function vanishes for $x \neq 0$! Although the name of $\delta(x)$ includes the word 'function', this object is not really a function because we cannot assign it a value at the origin. It is best considered to be the limit of a series of functions that all have unit area under their graphs but become more and more sharply peaked around the origin (see Figure 2.1).

The analogue of equation (1.31) is

$$\int \mathrm{d}x\, |\psi(x)|^2 = 1, \tag{2.43}$$

which expresses the physical requirement that there is unit probability of finding the particle at *some* value of x.

---

[4] The analogy would be clearer if we wrote $a(x)$ for $\psi(x)$, but for historical reasons the letter ψ is hard to avoid in this context.

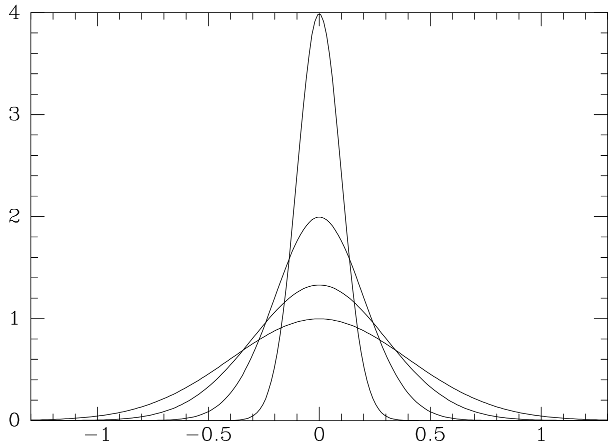


Figure 2.1 A series of Gaussians of unit area. The Dirac delta function is the limit of this series of functions as the dispersion tends to zero.

The analogue of equation (2.2) is

$$I = \int \mathrm{d}x\, |x\rangle\langle x|. \tag{2.44}$$

It is instructive to check that the operator that is defined by the right side of this equation really is the identity operator. Applying the operator to an arbitrary state $|\psi\rangle$ we find

$$I|\psi\rangle = \int \mathrm{d}x\, |x\rangle\langle x|\psi\rangle \tag{2.45}$$

By equations (2.37) and (2.38) the expression on the right of this equation is $|\psi\rangle$, so I is indeed the identity operator.

When we multiply (2.45) by $\langle\phi|$ on the left, we obtain an important formula

$$\langle\phi|\psi\rangle = \int \mathrm{d}x\, \langle\phi|x\rangle\langle x|\psi\rangle = \int \mathrm{d}x\, \phi^*(x)\psi(x), \tag{2.46}$$

where the second equality uses equations (2.38) and (2.39). Many practical problems reduce to the evaluation of an amplitude such as $\langle\phi|\psi\rangle$. The expression on the right of equation (2.46) is a well defined integral that evaluates to the desired number.

By analogy with equation (2.5), the **position operator** is

$$\hat{x} = \int \mathrm{d}x\, x|x\rangle\langle x|. \tag{2.47}$$

After applying $\hat{x}$ to a ket $|\psi\rangle$ we have a ket $|\phi\rangle = \hat{x}|\psi\rangle$ whose wavefunction $\phi(x') = \langle x'|\hat{x}|\psi\rangle$ is

$$\begin{aligned}\phi(x') = \langle x'|\hat{x}|\psi\rangle &= \int \mathrm{d}x\, x\langle x'|x\rangle\langle x|\psi\rangle \\ &= \int \mathrm{d}x\, x\delta(x-x')\psi(x) = x'\psi(x'),\end{aligned} \tag{2.48}$$

where we have used equations (2.38) and (2.40). Equation (2.48) states that the operator $\hat{x}$ simply multiplies the wavefunction $\psi(x)$ by its argument.

In the position representation, operators turn functions of x into other functions of x. An easy way of making a new function out of an old one is to differentiate it. So consider the operator $\hat{p}$ that is defined by

$$\langle x|\hat{p}|\psi\rangle = (\hat{p}\psi)(x) = -\mathrm{i}\hbar\frac{\partial\psi}{\partial x}. \tag{2.49}$$

Box 2.2: Proof that $\hat{p}$ is Hermitian

We have to show that for any states $|\phi\rangle$ and $|\psi\rangle$, $\langle\psi|\hat{p}|\phi\rangle = (\langle\phi|\hat{p}|\psi\rangle)^*$. We use equation (2.49) to write the left side of this equation in the position representation:

$$\langle\psi|\hat{p}|\phi\rangle = -\mathrm{i}\hbar\int \mathrm{d}x\,\psi^*(x)\frac{\partial\phi}{\partial x}.$$

Integrating by parts this becomes

$$\langle\psi|\hat{p}|\phi\rangle = -\mathrm{i}\hbar\left(\left[\psi^*\phi\right]_{-\infty}^{\infty} - \int \mathrm{d}x\,\phi(x)\frac{\partial\psi^*}{\partial x}\right).$$

We assume that all wavefunctions vanish at spatial infinity, so the term in square brackets vanishes, and

$$\langle\psi|\hat{p}|\phi\rangle = \mathrm{i}\hbar\int \mathrm{d}x\,\phi(x)\frac{\partial\psi^*}{\partial x} = (\langle\phi|\hat{p}|\psi\rangle)^*.$$

In Box 2.2 we show that the factor i ensures that $\hat{p}$ is a Hermitian operator. The factor $\hbar$ ensures that $\hat{p}$ has the dimensions of momentum:[5] we will find that $\hat{p}$ is the **momentum operator**. In Newtonian physics the momentum of a particle of mass m and velocity $\dot{x}$ is $m\dot{x}$, so let's use equation (2.34) to calculate $\mathrm{d}\langle\hat{x}\rangle/\mathrm{d}t$ and see whether it is $\langle\hat{p}\rangle/m$.

2.3.1 Hamiltonian of a particle

To calculate any time derivatives in quantum mechanics we need to know what the Hamiltonian operator H of our system is because H appears in the TDSE (2.26). Equation (2.5) defines H in the energy representation, but not how to write H in the position representation. We are going to have to make an informed guess and justify our guess later.

The Newtonian expression for the energy of a particle is

$$E = \tfrac{1}{2}m\dot{x}^2 + V = \frac{p^2}{2m} + V, \tag{2.50}$$

where $V(x)$ is the particle's potential energy. So we guess that the Hamiltonian of a particle is

$$H = \frac{\hat{p}^2}{2m} + V(\hat{x}), \tag{2.51}$$

where the square of $\hat{p}$ means the act of operating with $\hat{p}$ twice ($\hat{p}^2 \equiv \hat{p}\hat{p}$). The meaning of $V(\hat{x})$ is given by equation (2.20) with V and $\hat{x}$ substituted for f and R. Working from that equation in close analogy with the calculation in equation (2.48) demonstrates that in the position representation the operator $V(\hat{x})$ acts on a wavefunction $\psi(x)$ simply by multiplying ψ by $V(x)$. That is, $\langle x|V(\hat{x})|\psi\rangle = V(x)\psi(x)$.

Now that we have guessed that H is given by equation (2.51), the next step in the calculation of the rate of change of $\langle\hat{x}\rangle$ is to evaluate the commutator of $\hat{x}$ and H. Making use of equations (2.22) we find

$$\begin{aligned}[\hat{x},H] &= \left[\hat{x},\frac{\hat{p}^2}{2m}+V\right] = \frac{[\hat{x},\hat{p}\hat{p}]}{2m} + [\hat{x},V(\hat{x})]\\ &= \frac{[\hat{x},\hat{p}]\hat{p}+\hat{p}[\hat{x},\hat{p}]}{2m}.\end{aligned} \tag{2.52}$$

---

[5] Planck's constant $h = 2\pi\hbar$ has dimensions of distance × momentum, or, equivalently, energy × time, or, most simply, angular momentum.

In the last equality we have used the fact that $[\hat{x}, V(\hat{x})] = 0$, which follows because both $\hat{x}$ and $V(\hat{x})$ act by multiplication, and ordinary multiplication is a commutative operation. We now have to determine the value of the commutator $[\hat{x}, \hat{p}]$. We return to the definition (2.49) of $\hat{p}$ and calculate the wavefunction produced by applying $[\hat{x}, \hat{p}]$ to an arbitrary state $|\psi\rangle$

$$\begin{aligned}[\hat{x}, \hat{p}]\psi(x) = \langle x|(\hat{x}\hat{p} - \hat{p}\hat{x})|\psi\rangle &= -\mathrm{i}\hbar\left(x\frac{\partial\psi}{\partial x} - \frac{\partial(x\psi)}{\partial x}\right) \\ &= \mathrm{i}\hbar\psi(x).\end{aligned} \tag{2.53}$$

Since this equation holds for any ψ, we have the operator equation

$$[\hat{x}, \hat{p}] = \mathrm{i}\hbar. \tag{2.54}$$

This key result, that the commutator of $\hat{x}$ with $\hat{p}$ is the constant $\mathrm{i}\hbar$, is called a **canonical commutation relation**.[6] Two observables whose commutator is $\pm\mathrm{i}\hbar$ are said to be **canonically conjugate** to one another, or **conjugate observables**.

Finally we have the hoped-for relation between p and $\dot{x}$: substituting equations (2.53) and (2.54) into equation (2.34) we have

$$\begin{aligned}\frac{\mathrm{d}\,\langle x\rangle}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\langle\psi|\hat{x}|\psi\rangle = -\frac{\mathrm{i}}{\hbar}\langle\psi|[\hat{x}, H]|\psi\rangle = -\frac{\mathrm{i}}{\hbar}\frac{\mathrm{i}\hbar}{m}\langle\psi|\hat{p}|\psi\rangle \\ &= \frac{1}{m}\langle\hat{p}\rangle\,.\end{aligned} \tag{2.55}$$

This result makes it highly plausible that $\hat{p}$ is indeed the momentum operator.

A calculation of the rate of change of $\langle\hat{p}\rangle$ will increase the plausibility still further. Again working from (2.34) and using (2.51) we have

$$\frac{\mathrm{d}\,\langle\hat{p}\rangle}{\mathrm{d}t} = -\frac{\mathrm{i}}{\hbar}\langle[\hat{p}, H]\rangle = -\frac{\mathrm{i}}{\hbar}\langle[\hat{p}, V]\rangle\,. \tag{2.56}$$

Since $[\hat{p}, \hat{x}] = -\mathrm{i}\hbar$ is just a number, equation (2.25) for the commutator of one operator with a function of another operator can be used to evaluate $[\hat{p}, V(\hat{x})]$. We then have

$$\frac{\mathrm{d}\,\langle\hat{p}\rangle}{\mathrm{d}t} = -\left\langle\frac{\mathrm{d}V}{\mathrm{d}\hat{x}}\right\rangle. \tag{2.57}$$

That is, the expectation of the rate of change of the momentum is equal to the expectation of the force on the particle. Thus we have recovered Newton's second law from the TDSE. This achievement gives us confidence that (2.51) is the correct expression for H.

2.3.2 Wavefunction for well defined momentum

From the discussion below equation (2.11) we know that the state $|p\rangle$ in which a measurement of the momentum will certainly yield the value p has to be an eigenstate of $\hat{p}$. We find the wavefunction $u_p(x) = \langle x|p\rangle$ of this important state by using equation (2.49) to write the defining equation $\hat{p}|p\rangle = p|p\rangle$ in the position representation:

$$\langle x|\hat{p}|p\rangle = -\mathrm{i}\hbar\frac{\partial u_p}{\partial x} = p\langle x|p\rangle = pu_p(x). \tag{2.58}$$

The solution of this differential equation is

$$u_p(x) = A\mathrm{e}^{\mathrm{i}px/\hbar}. \tag{2.59}$$

---

[6] The name that derives from 'canonical coordinates' in Hamilton's formulation of classical mechanics.

Box 2.3: Gaussian integrals

Consider the integral

$$I \equiv \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-(b^2x^2+ax)}, \tag{1}$$

where a and b are constants. We observe that $b^2x^2 + ax = (bx + a/2b)^2 - a^2/4b^2$. Thus we may write $I = \mathrm{e}^{a^2/4b^2} b^{-1} \int \mathrm{d}z\, \mathrm{e}^{-z^2}$, where $z \equiv bx + a/2b$. The integral is equal to $\sqrt{\pi}$. Hence we have the very useful result

$$\int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-(b^2x^2+ax)} = \frac{\sqrt{\pi}}{b}\, \mathrm{e}^{a^2/4b^2}. \tag{2}$$

Hence the wavefunction of a particle of well defined momentum is a plane wave with wavelength $\lambda = 2\pi/k = h/\sqrt{2mE}$, where m is the particle's mass and E its kinetic energy; λ is called the particle's **de Broglie wavelength**.[7]

If we try to choose the constant A in (2.59) to ensure that u_p satisfies the usual normalisation condition (2.43), we will fail because the integral over all x of $|\mathrm{e}^{\mathrm{i}px/\hbar}|^2 = 1$ is undefined. Instead we choose A as follows. By analogy with (2.40) we require $\langle p'|p\rangle = \delta(p - p')$. When we use (2.44) to insert an identity operator into this expression, it becomes

$$\delta(p-p') = \int \mathrm{d}x\, \langle p'|x\rangle\langle x|p\rangle = |A|^2 \int \mathrm{d}x\, \mathrm{e}^{\mathrm{i}(p-p')x/\hbar} = 2\pi\hbar|A|^2\delta(p-p'), \tag{2.60}$$

where we have used equation (B.12) to evaluate the integral. Thus $|A|^2 = h^{-1}$, where $h = 2\pi\hbar$ is Planck's constant, and the correctly normalised wavefunction of a particle of momentum p is

$$u_p(x) \equiv \langle x|p\rangle = \frac{1}{\sqrt{h}}\, \mathrm{e}^{\mathrm{i}px/\hbar}. \tag{2.61}$$

The uncertainty principle It follows from (2.61) that the position of a particle that has well defined momentum is maximally uncertain: all values of x are equally probable because $|u_p|^2$ is independent of x. This phenomenon is said to be a consequence of the **uncertainty principle**, namely that when an observable has a well-defined value, all values of the canonically conjugate observable are equally probable.

We can gain useful insight into the workings of the uncertainty principle by calculating the variance in momentum measurements for states in which measurements of position are subject to varying degrees of uncertainty. For definiteness we take the probability density $|\psi(x)|^2$ to be a Gaussian distribution of dispersion σ. So we write

$$\psi(x) = \frac{1}{(2\pi\sigma^2)^{1/4}}\, \mathrm{e}^{-x^2/4\sigma^2}. \tag{2.62}$$

With equations (2.46) and (2.61) we find that in this state the amplitude to measure momentum p is

$$\langle p|\psi\rangle = \int \mathrm{d}x\, u_p^*(x)\psi(x) = \frac{1}{\sqrt{h}(2\pi\sigma^2)^{1/4}} \int \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}px/\hbar}\, \mathrm{e}^{-x^2/4\sigma^2}. \tag{2.63}$$

---

[7] Louis de Broglie (1892–1987) was the second son of the Duc de Broglie. In 1924 his PhD thesis introduced the concept of matter waves, by considering relativistic invariance of phase. For this work he won the 1929 Nobel prize for physics. In later years he struggled to find a causal rather than probabilistic interpretation of quantum mechanics.

Box 2.3 explains how integrals of this type are evaluated. Setting $a = \mathrm{i}p/\hbar$ and $b = (2\sigma)^{-1}$ in equation (2) of Box 2.3 we find

$$\langle p|\psi\rangle = \frac{2\sigma\sqrt{\pi}}{\sqrt{h}(2\pi\sigma^2)^{1/4}}\,\mathrm{e}^{-\sigma^2 p^2/\hbar^2} = \frac{1}{(2\pi\hbar^2/4\sigma^2)^{1/4}}\,\mathrm{e}^{-\sigma^2 p^2/\hbar^2}. \tag{2.64}$$

The probability density $|\langle p|\psi\rangle|^2$ is a Gaussian centred on zero with a dispersion σ_p in p that equals $\hbar/2\sigma$. Thus, the more sharply peaked the particle's probability distribution is in x, the broader the distribution is in p. The product of the dispersions in x and p is always $\frac{1}{2}\hbar$: $\sigma_p\sigma = \frac{1}{2}\hbar$.

This trade-off between the uncertainties in x and p arises because when we expand $|\psi\rangle$ in eigenkets of $\hat{p}$, localisation of the probability amplitude $\psi(x)$ is caused by interference between states of different momenta: in the position representation, these states are plane waves of wavelength h/p that have the same amplitude everywhere, and interference between waves of very different wavelengths is required if the region of constructive interference is to be strongly confined.

2.3.3 Dynamics of a free particle

We now consider the motion of a free particle – one that is subject to no forces so we can drop the potential term in the Hamiltonian (2.51). Consequently, the Hamiltonian of a free particle,

$$H = \frac{\hat{p}^2}{2m}, \tag{2.65}$$

is a function of $\hat{p}$ alone, so its eigenkets will be the eigenkets (2.61) of $\hat{p}$. By expressing any ket $|\psi\rangle$ as a linear combination of these eigenkets, and using the basic time-evolution equation (2.32), we can follow the motion of a particle from the initial state $|\psi\rangle$. We illustrate this procedure with the case in which ψ corresponds to the particle being approximately at the origin with momentum near some value p_0. Equation (2.64) gives $\langle p|\psi\rangle$ for the case in which p_0 vanishes. The amplitude distribution that we require is

$$\langle p|\psi,0\rangle = \frac{1}{(2\pi\hbar^2/4\sigma^2)^{1/4}}\,\mathrm{e}^{-\sigma^2(p-p_0)^2/\hbar^2}. \tag{2.66}$$

We can now use (2.32) to obtain the wavefunction t units of time later

$$\begin{aligned}\langle x|\psi,t\rangle &= \int \mathrm{d}p\,\langle x|p\rangle\langle p|\psi,0\rangle \mathrm{e}^{-\mathrm{i}p^2t/2m\hbar}\\ &= \frac{1}{\sqrt{h}(2\pi\hbar^2/4\sigma^2)^{1/4}}\int \mathrm{d}p\,\mathrm{e}^{\mathrm{i}px/\hbar}\mathrm{e}^{-\sigma^2(p-p_0)^2/\hbar^2}\mathrm{e}^{-\mathrm{i}p^2t/2m\hbar}.\end{aligned} \tag{2.67}$$

Evaluating the integral in this expression involves some tiresome algebra – you can find the details in Box 2.4 if you are interested. We want the probability density at time t of finding the particle at x, which is the mod-square of equation (2.67). From the last equation of Box 2.4 we have

$$|\langle x|\psi,t\rangle|^2 = \frac{\sigma}{\sqrt{2\pi}\hbar^2|b|^2}\exp\left\{\frac{-(x-p_0t/m)^2\sigma^2}{2\hbar^4|b|^4}\right\}. \tag{2.68}$$

This is a Gaussian distribution whose centre moves with the velocity p_0/m associated with the most probable momentum in the initial data (2.66). The square of the Gaussian's dispersion is

$$\sigma^2(t) = \sigma^2 + \left(\frac{\hbar t}{2m\sigma}\right)^2. \tag{2.69}$$

Box 2.4: Evaluating the integral in equation (2.67)

The integral is of the form discussed in Box 2.3. To clean it up we replace the p^2 in the third exponential with $(p-p_0)^2+2p_0p-p_0^2$ and gather together all three exponents:

$$\begin{aligned}\langle x|\psi,t\rangle =& \frac{\mathrm{e}^{\mathrm{i}p_0^2t/2m\hbar}}{\sqrt{\hbar}(2\pi\hbar^2/4\sigma^2)^{1/4}}\\ &\times\int \mathrm{d}p\,\exp\left\{\frac{\mathrm{i}p}{\hbar}\left(x-\frac{p_0t}{m}\right)-(p-p_0)^2\left(\frac{\sigma^2}{\hbar^2}+\frac{\mathrm{i}t}{2m\hbar}\right)\right\}.\end{aligned}$$

In Box 2.3 we now set

$$a=\frac{\mathrm{i}}{\hbar}\left(x-\frac{p_0t}{m}\right)\quad ;\quad b^2=\left(\frac{\sigma^2}{\hbar^2}+\frac{\mathrm{i}t}{2m\hbar}\right)$$

and conclude that

$$\langle x|\psi,t\rangle=\frac{\mathrm{e}^{\mathrm{i}p_0^2t/2m\hbar}}{\sqrt{\hbar}(2\pi\hbar^2/4\sigma^2)^{1/4}}\exp\left\{\frac{\mathrm{i}p_0}{\hbar}\left(x-\frac{p_0t}{m}\right)\right\}\frac{\sqrt{\pi}}{b}\mathrm{e}^{-(x-p_0t/m)^2/4\hbar^2b^2}.$$

This is a complicated result because b is a complex number, but its mod-square, equation (2.68), is relatively simple.

We saw above that in the initial data the uncertainty in p is $\sim\sigma_p=\hbar/2\sigma$, which translates to an uncertainty in velocity $\Delta_v\sim\hbar/2m\sigma$. After time t this uncertainty should lead to an additional uncertainty in position $\Delta_x=\Delta_vt\sim\hbar t/2m\sigma$ in perfect agreement with equation (2.69).

These results complete the demonstration that the identification of the operator $\hat{p}$ defined by equation (2.49) with the momentum operator, together with the Hamiltonian (2.51), enable us to recover as much of Newtonian mechanics as we expect to continue valid outside the classical regime. The idea that in an appropriate limit the predictions of quantum mechanics should agree with classical mechanics is often called the **correspondence principle**. The discipline of checking that one's calculations comply with the correspondence principle is useful in several ways: (i) it provides a check on the calculations, helping one to locate missing factors of i or incorrect signs, (ii) it deepens one's understanding of classical physics, and (iii) it draws attention to novel predictions of quantum mechanics that have no counterparts in classical mechanics.

In the process of checking the correspondence principle for a free particle we have stumbled on a new principle, the uncertainty principle, which implies that the more tightly constrained the value of one observable is, the more uncertain the value of the conjugate variable must be. Notice that these uncertainties do not arise from measurement errors: we have assumed that x and p can be measured exactly. The uncertainties we have discussed are inherent in the situation and can only be increased by deficiencies in the measurement process.

Our calculations have also shown how far-reaching the principle of quantum interference is: equation (2.67), upon which our understanding of the dynamics of a free particle rests, expresses the amplitude for the particle to be found at x at time t as an integral over momenta of the amplitude to travel at momentum p. It is through interference between the infinite number of contributing amplitudes that classically recognisable dynamics is recovered. Had we mod-squared the amplitudes before adding them, as classical probability theory would suggest, we would have obtained entirely unphysical results.

2.3.4 Back to two-slit interference

When we discussed the two-slit interference experiment in §1.2.1, we stated without proof that $\phi_1 - \phi_2 \propto x$, where $\phi_i(x)$ is the phase of the amplitude $A_i(x)$ for an electron to arrive at the point x on the screen P after passing through the slit S_i. We can now justify this assertion and derive the constant of proportionality. Once the constant has been determined, it is possible to assess the feasibility of the experiment from a practical point of view.

We assume that the quantum state of an electron as it emerges from the electron gun can be approximated by a state of well defined momentum $|p\rangle$. So the wavefunction between the gun and the screen with the slits is a plane wave of wavelength $\lambda = h/p$. As an electron passes through a slit we assume that it is deflected slightly but retains its former kinetic energy. So we approximate its wavefunction after passing through the slit by a wave that is no longer plane, but still has wavelength λ. Hence the phase of this wave at position x on the screen P will be the phase at the slit plus $2\pi D/\lambda = pD/\hbar$, where $D(x)$ is the distance from x to the slit. By Pythagoras's theorem

$$D = \sqrt{L^2 + (x \pm s)^2}, \tag{2.70}$$

where L is the distance between the screen with the slits and P, $2s$ is the distance between the slits, and the plus sign applies for one slit and the minus sign to the other. We assume that both x and s are much smaller than L so the square root can be expanded by the binomial theorem. We then find that the difference of the phases is

$$\phi_1 - \phi_2 \simeq \frac{2psx}{\hbar L}. \tag{2.71}$$

The distance X between the dark bands on P is the value of x for which the left side becomes 2π, so

$$X = \frac{hL}{2ps}. \tag{2.72}$$

Let's put some numbers into this formula. Since $h = 6.63 \times 10^{-34}\,\mathrm{J\,s}$ is very small, there is a danger that X will come out too small to produce observable bands. Therefore we choose L fairly large and both p and s small. Suppose we adopt 1 m for L and 1 μm for s. From the Hamiltonian (2.65) we have $p = \sqrt{2mE}$. A reasonable energy for the electrons is $E = 100\,\mathrm{eV} = 1.6 \times 10^{-17}\,\mathrm{J}$, which yields $p = 5.5 \times 10^{-24}\,\mathrm{Ns}$, and $X = 0.057\,\mathrm{mm}$. Hence there should be no difficulty observing a sinusoidal pattern that has this period.

What do the numbers look like for bullets? On a firing range we can probably stretch L to $1000\,\mathrm{m}$. The distance between the slits clearly has to be larger than the diameter of a bullet, so we take $s = 1\,\mathrm{cm}$. A bullet weighs $\sim 10\,\mathrm{gm}$ and travels at $\sim 300\,\mathrm{m\,s^{-1}}$. Equation (2.72) now yields $X \sim 10^{-29}\,\mathrm{m}$. So it is not surprising that fire-arms manufacturers find classical mechanics entirely satisfactory.

2.3.5 Generalisation to three dimensions

Real particles move in three dimensions rather than one. Fortunately, the generalisation to three dimensions of what we have done in one dimension is straightforward.

The x, y and z coordinates of a particle are three distinct observables. Their operators commute with one another:

$$[\hat{x}_i, \hat{x}_j] = 0. \tag{2.73}$$

Since these are commuting observables, there is a complete set of mutual eigenkets, $\{|\mathbf{x}\rangle\}$. We can express any state of the system, $|\psi\rangle$, as a linear combination of these kets:

$$|\psi\rangle = \int \mathrm{d}^3\mathbf{x}\, \langle \mathbf{x}|\psi\rangle\, |\mathbf{x}\rangle = \int \mathrm{d}^3\mathbf{x}\, \psi(\mathbf{x})\, |\mathbf{x}\rangle, \tag{2.74}$$

where the wavefunction $\psi(\mathbf{x})$ is now a function of three variables, and the integral is over all space.

The x, y and z components of the particle's momentum $\hat{\mathbf{p}}$ commute with one another:

$$[\hat{p}_i, \hat{p}_j] = 0. \tag{2.75}$$

In the position representation, these operators are represented by partial derivatives with respect to their respective coordinates

$$\hat{p}_i = -\mathrm{i}\hbar\frac{\partial}{\partial x_i} \quad \text{so} \quad \hat{\mathbf{p}} = -\mathrm{i}\hbar\nabla\ . \tag{2.76}$$

The momenta commute with all operators except their conjugate coordinate, so the canonical commutation relations are

$$[\hat{x}_i, \hat{p}_j] = \mathrm{i}\hbar\delta_{ij}. \tag{2.77}$$

In §4.2 we will understand the origin of the factor δ_{ij}. Since the three momentum operators commute with one another, there is a complete set of mutual eigenstates. Analogously to equation (2.61), the wavefunction of the state with well defined momentum $\mathbf{p}$ is

$$\langle\mathbf{x}|\mathbf{p}\rangle = \frac{1}{h^{3/2}}\mathrm{e}^{\mathrm{i}\mathbf{x}\cdot\mathbf{p}/\hbar}. \tag{2.78}$$

The virial theorem We illustrate the use of these relations by deriving a relation between the kinetic and potential energies of a particle that is in a stationary state. In §2.2.1 we showed that all expectation values are time-independent when a system is in a stationary state. We apply this result to the operator $\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}$

$$\begin{aligned}0 = \mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\langle\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}\rangle &= \langle E|\Big[\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}, \frac{\hat{p}^2}{2m} + V(\hat{\mathbf{x}})\Big]|E\rangle \\ &= \frac{1}{2m}\langle E|[\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}, \hat{p}^2]|E\rangle + \langle E|[\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}, V(\hat{\mathbf{x}})]|E\rangle.\end{aligned} \tag{2.79}$$

The first commutator can be expanded thus

$$[\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}, \hat{p}^2] = \sum_{jk}[\hat{x}_j\hat{p}_j, \hat{p}_k^2] = \sum_{jk}[\hat{x}_j, \hat{p}_k^2]\hat{p}_j = \sum_{jk} 2\mathrm{i}\hbar\hat{p}_k\delta_{jk}\hat{p}_j = 2\mathrm{i}\hbar\hat{p}^2. \tag{2.80}$$

In the position representation the second commutator is simply

$$[\hat{\mathbf{x}}\cdot\hat{\mathbf{p}}, V(\hat{\mathbf{x}})] = -\mathrm{i}\hbar\mathbf{x}\cdot\nabla V(\mathbf{x}). \tag{2.81}$$

When we put these results back into (2.79) and rearrange, we obtain the **virial theorem**

$$2\langle E|\frac{\hat{p}^2}{2m}|E\rangle = \langle E|(\mathbf{x}\cdot\nabla V)|E\rangle. \tag{2.82}$$

In important applications the potential is proportional to some power of distance from the origin: $V(\mathbf{x}) = C|\mathbf{x}|^\alpha$. Then, because $\nabla|\mathbf{x}| = \mathbf{x}/|\mathbf{x}|$, the operator on the right is $\mathbf{x}\cdot\nabla V = \alpha C|\mathbf{x}|^\alpha = \alpha V$ and the virial theorem becomes

$$2\langle E|\frac{\hat{p}^2}{2m}|E\rangle = \alpha\langle E|V|E\rangle. \tag{2.83}$$

So twice the kinetic energy is equal to α times the potential energy. For example, for a harmonic oscillator $\alpha = 2$, so kinetic and potential energies are equal. The other important example is motion in an inverse-square force field, such as the electrostatic field of an atomic nucleus. In this case $\alpha = -1$, so twice the kinetic energy plus the potential energy vanishes. Equivalently, the kinetic energy is equal in magnitude but opposite in sign to the total energy.

Problems

2.1 How is a wave-function $\psi(x)$ written in Dirac's notation? What's the physical significance of the complex number $\psi(x)$ for given x?

2.2 Let Q be an operator. Under what circumstances is the complex number $\langle a|Q|b\rangle$ equal to the complex number $(\langle b|Q|a\rangle)^*$ for any states $|a\rangle$ and $|b\rangle$?

2.3 Let Q be the operator of an observable and let $|\psi\rangle$ be the state of our system.

a. What are the physical interpretations of $\langle\psi|Q|\psi\rangle$ and $|\langle q_n|\psi\rangle|^2$, where $|q_n\rangle$ is the n^{th} eigenket of the observable Q and q_n is the corresponding eigenvalue?

b. What is the operator $\sum_n |q_n\rangle\langle q_n|$, where the sum is over all eigenkets of Q? What is the operator $\sum_n q_n|q_n\rangle\langle q_n|$?

c. If $u_n(x)$ is the wavefunction of the state $|q_n\rangle$, write dow an integral that evaluates to $\langle q_n|\psi\rangle$.

2.4 What does it mean to say that two operators commute? What is the significance of two observables having mutually commuting operators?

Given that the commutator $[P, Q] \neq 0$ for some observables P and Q, does it follow that for all $|\psi\rangle \neq 0$ we have $[P, Q]|\psi\rangle \neq 0$?

2.5 Let $\psi(x, t)$ be the correctly normalised wavefunction of a particle of mass m and potential energy $V(x)$. Write down expressions for the expectation values of (a) x; (b) x^2; (c) the momentum p_x; (d) p_x^2; (e) the energy.

What is the probability that the particle will be found in the interval (x_1, x_2)?

2.6 Write down the time-independent (TISE) and the time-dependent (TDSE) Schrödinger equations. Is it necessary for the wavefunction of a system to satisfy the TDSE? Under what circumstances does the wavefunction of a system satisfy the TISE?

2.7 Why is the TDSE first-order in time, rather than second-order like Newton's equations of motion?

2.8 A particle is confined in a potential well such that its allowed energies are $E_n = n^2\mathcal{E}$, where $n = 1, 2, \ldots$ is an integer and $\mathcal{E}$ a positive constant. The corresponding energy eigenstates are $|1\rangle$, $|2\rangle$, …, $|n\rangle$, …. At $t = 0$ the particle is in the state

$$|\psi(0)\rangle = 0.2|1\rangle + 0.3|2\rangle + 0.4|3\rangle + 0.843|4\rangle. \tag{2.84}$$

a. What is the probability, if the energy is measured at $t = 0$ of finding a number smaller than $6\mathcal{E}$?

b. What is the mean value and what is the rms deviation of the energy of the particle in the state $|\psi(0)\rangle$?

c. Calculate the state vector $|\psi\rangle$ at time t. Do the results found in (a) and (b) for time t remain valid for arbitrary time t?

d. When the energy is measured it turns out to be $16\mathcal{E}$. After the measurement, what is the state of the system? What result is obtained if the energy is measured again?

2.9 A system has a time-independent Hamiltonian that has spectrum $\{E_n\}$. Prove that the probability P_k that a measurement of energy will yield the value E_k is is time-independent. Hint: you can do this either from Ehrenfest's theorem, or by differentiating $\langle E_k|\psi\rangle$ w.r.t. t and using the TDSE.

2.10 Let $\psi(x)$ be a properly normalised wavefunction and Q an operator on wavefunctions. Let $\{q_r\}$ be the spectrum of Q and $\{u_r(x)\}$ be the corresponding correctly normalised eigenfunctions. Write down an expression for the probability that a measurement of Q will yield the value q_r. Show that $\sum_r P(q_r|\psi) = 1$. Show further that the expectation of Q is $\langle Q\rangle \equiv \int_{-\infty}^{\infty} \psi^* \hat{Q}\psi \,\mathrm{d}x$.[8]

2.11 Find the energy of neutron, electron and electromagnetic waves of wavelength 0.1 nm.

2.12 Neutrons are emitted from an atomic pile with a Maxwellian distribution of velocities for temperature 400 K. Find the most probable de Broglie wavelength in the beam.

2.13 A beam of neutrons with energy E runs horizontally into a crystal. The crystal transmits half the neutrons and deflects the other half vertically upwards. After climbing to height H these neutrons are deflected through 90° onto a horizontal path parallel to the originally transmitted beam. The two horizontal beams now move a distance L down the laboratory, one distance H above the other. After going distance L, the lower beam is deflected vertically upwards and is finally deflected into the path of the upper beam such that the two beams are co-spatial as they enter the detector. Given that particles in both the lower and upper beams are in states of well-defined momentum, show that the wavenumbers k, k' of the lower and upper beams are related by

$$k' \simeq k\left(1 - \frac{m_\mathrm{n} g H}{2E}\right). \tag{2.85}$$

In an actual experiment (R. Colella et al., 1975, Phys. Rev. Let., 34, 1472) $E = 0.042\,\mathrm{eV}$ and $LH \sim 10^{-3}\,\mathrm{m}^2$ (the actual geometry was slightly different). Determine the phase difference between the two beams at the detector. Sketch the intensity in the detector as a function of H.

2.14 A particle moves in the potential $V(\mathbf{x})$ and is known to have energy E_n. (a) Can it have well defined momentum for some particular $V(\mathbf{x})$? (b) Can the particle simultaneously have well-defined energy and position?

2.15 The states $\{|1\rangle, |2\rangle\}$ form a complete orthonormal set of states for a two-state system. With respect to these basis states the operator σ_y has matrix

$$\sigma_y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}. \tag{2.86}$$

Could σ be an observable? What are its eigenvalues and eigenvectors in the $\{|1\rangle, |2\rangle\}$ basis? Determine the result of operating with σ_y on the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle). \tag{2.87}$$

2.16 A three-state system has a complete orthonormal set of states $|1\rangle, |2\rangle, |3\rangle$. With respect to this basis the operators H and B have matrices

$$H = \hbar\omega \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \qquad B = b \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{2.88}$$

where ω and b are real constants.

a. Are H and B Hermitian?

b. Write down the eigenvalues of H and find the eigenvalues of B. Solve for the eigenvectors of both H and B. Explain why neither matrix uniquely specifies its eigenvectors.

c. Show that H and B commute. Give a basis of eigenvectors common to H and B.

---

[8] In the most elegant formulation of qantum mechanics, this last result is the basic postulate of the theory, and one *derives* other rules for the physical interpretation of the q_n, a_n etc. from it – see J. von Neumann, *Mathematical Foundations of Quantum Mechanics.*

2.17 Given that A and B are Hermitian operators, show that $\mathrm{i}[A,B]$ is a Hermitian operator.

2.18 Given a ordinary function $f(x)$ and an operator R, the operator $f(R)$ is defined to be

$$f(R) = \sum_i f(r_i)|r_i\rangle\langle r_i|, \tag{2.89}$$

where r_i are the eigenvalues of R and $|r_i\rangle$ are the associated eigenkets. Show that when $f(x) = x^2$ this definition implies that $f(R) = RR$, that is, that operating with $f(R)$ is equivalent to applying the operator R twice. What bearing does this result have in the meaning of e^R?

2.19 Show that if there is a complete set of mutual eigenkets of the Hermitian operators A and B, then $[A,B] = 0$. Explain the physical significance of this result.

2.20 Given that for any two operators $(AB)^\dagger = B^\dagger A^\dagger$, show that

$$(ABCD)^\dagger = D^\dagger C^\dagger B^\dagger A^\dagger. \tag{2.90}$$

2.21 Prove for any four operators A, B, C, D that

$$[ABC, D] = AB[C,D] + A[B,D]C + [A,D]BC. \tag{2.91}$$

Explain the similarity with the rule for differentiating a product.

2.22 Show that for any three operators A, B and C, the **Jacobi identity** holds:

$$[A,[B,C]] + [B,[C,A]] + [C,[A,B]] = 0. \tag{2.92}$$

2.23 Show that a classical harmonic oscillator satisfies the virial equation $2\langle\mathrm{KE}\rangle = \alpha\langle\mathrm{PE}\rangle$ and determine the relevant value of α.

2.24 A classical fluid of density $\rho(\mathbf{x})$ flows with velocity $\mathbf{v}(\mathbf{x})$. By differentiating with respect to time the mass $m \equiv \int_V \mathrm{d}^3\mathbf{x}\,\rho$ contained in an arbitrary volume V, show that conservation of mass requires that

$$\frac{\partial\rho}{\partial t} + \boldsymbol{\nabla}\cdot(\rho\mathbf{v}) = 0. \tag{2.93}$$

Hint: the flux of matter at any point is $\rho\mathbf{v}$ and the integral of this flux over the boundary of V must equal the rate of accumulation of mass within V.

$\mathbf{J}$ is defined to be

$$\mathbf{J}(\mathbf{x}) \equiv \frac{\mathrm{i}\hbar}{2m}\Big(\psi\boldsymbol{\nabla}\psi^* - \psi^*\boldsymbol{\nabla}\psi\Big), \tag{2.94}$$

where $\psi(\mathbf{x})$ is the wavefunction of a spinless particle of mass m. Working from the TDSE, show that

$$\frac{\partial|\psi|^2}{\partial t} + \boldsymbol{\nabla}\cdot\mathbf{J} = 0. \tag{2.95}$$

Give a physical interpretation of this result.

Show that when we write the wavefunction in amplitude-modulus form, $\psi = |\psi|\mathrm{e}^{\mathrm{i}\theta}$,

$$\mathbf{J} = |\psi|^2\frac{\hbar\boldsymbol{\nabla}\theta}{m}. \tag{2.96}$$

Interpret this result physically. Given that $\psi = A\mathrm{e}^{\mathrm{i}(kz-\omega t)} + B\mathrm{e}^{-\mathrm{i}(kz+\omega t)}$, where A and B are constants, show that

$$\mathbf{J} = v\big(|A|^2 - |B|^2\big)\,\hat{\mathbf{z}}, \tag{2.97}$$

where $v = \hbar k/m$. Interpret the result physically.

3

Harmonic oscillators and magnetic fields

Harmonic oscillators are of enormous importance for physics because most of condensed-matter physics and quantum electrodynamics centre on weakly perturbed harmonic oscillators. The reason harmonic oscillators are so common is simple. The points of equilibrium of a particle that moves in a potential $V(x)$ are points at which the force $-\mathrm{d}V/\mathrm{d}x$ vanishes. When we place the origin of x at such a point, the Maclaurin expansion of V becomes $V(x) = \text{constant} + \frac{1}{2}V''x^2 + \mathrm{O}(x^3)$, and the force on the particle becomes $F = -V''x + \mathrm{O}(x^2)$. Consequently, for sufficiently small excursions from the point of equilibrium, the particle's motion will be well approximated by a harmonic oscillator.

Besides providing the background to a great many branches of physics, our analysis of a harmonic oscillator will introduce a technique that we will use twice more in our analysis of the hydrogen atom. As a bonus, we will find that our results for the harmonic oscillator enable us to solve another important, and apparently unrelated problem: the motion of a charged particle in a uniform magnetic field.

3.1 Stationary states of a harmonic oscillator

We can build a harmonic oscillator by placing a particle in a potential that increases quadratically with distance from the origin. Hence an appropriate Hamiltonian is given by equation (2.51) with $V \propto x^2$.[1] For later convenience we choose the constant of proportionality such that H becomes

$$H = \frac{1}{2m}\{p^2 + (m\omega x)^2\}. \tag{3.1}$$

In §2.2 we saw that the dynamical evolution of a system follows immediately once we know the eigenvalues and eigenkets of H. So we now determine these quantities for the Hamiltonian (3.1).

---

[1] In the last chapter we distinguished the position and momentum operators from their eigenvalues with hats. Henceforth we drop the hats; the distinction between operator and eigenvalue should be clear from the context.

We next introduce the dimensionless operator

$$A \equiv \frac{m\omega x + \mathrm{i}p}{\sqrt{2m\hbar\omega}}. \tag{3.2a}$$

This operator isn't Hermitian. Bearing in mind that x and p *are* Hermitian, from the rules in Table 2.1 we see that its adjoint is

$$A^\dagger = \frac{m\omega x - \mathrm{i}p}{\sqrt{2m\hbar\omega}}. \tag{3.2b}$$

The product $A^\dagger A$ is

$$A^\dagger A = \frac{1}{2m\hbar\omega}\left\{(m\omega x)^2 + \mathrm{i}m\omega[x,p] + p^2\right\} = \frac{H}{\hbar\omega} - \tfrac{1}{2}, \tag{3.3}$$

where we have used the canonical commutation relation (2.54). This equation can be rewritten $H/(\hbar\omega) = A^\dagger A + \frac{1}{2}$, so A is rather nearly the square root of the dimensionless Hamiltonian $H/\hbar\omega$.

The commutator of A and $A^\dagger$ is

$$\begin{aligned}
[A^\dagger, A] &= \frac{1}{2m\hbar\omega}[m\omega x - \mathrm{i}p, m\omega x + \mathrm{i}p] \\
&= \frac{1}{2m\hbar\omega}\left\{m^2\omega^2[x,x] + \mathrm{i}m\omega\big([x,p] - [p,x]\big) + [p,p]\right\} \\
&= -1,
\end{aligned} \tag{3.4}$$

where we have exploited the rules of equations (2.22).

We now multiply by $A^\dagger$ both sides of the defining relation of $|E_n\rangle$, namely $H|E_n\rangle = E_n|E_n\rangle$:

$$A^\dagger E_n|E_n\rangle = A^\dagger H|E_n\rangle = \big(HA^\dagger + [A^\dagger, H]\big)|E_n\rangle. \tag{3.5}$$

We use equation (3.3) to replace the H in the commutator by $\hbar\omega(A^\dagger A + \frac{1}{2})$ and then use equation (3.4):

$$\begin{aligned}
E_n A^\dagger|E_n\rangle &= \big(HA^\dagger + \hbar\omega[A^\dagger, A^\dagger A + \tfrac{1}{2}]\big)|E_n\rangle = \big(HA^\dagger + \hbar\omega A^\dagger[A^\dagger, A]\big)|E_n\rangle \\
&= \big(HA^\dagger - \hbar\omega A^\dagger\big)|E_n\rangle.
\end{aligned} \tag{3.6}$$

We have shown that

$$(E_n + \hbar\omega)\big(A^\dagger|E_n\rangle\big) = H\big(A^\dagger|E_n\rangle\big). \tag{3.7}$$

Provided $|b\rangle \equiv A^\dagger|E_n\rangle$ has non-zero length-squared, this shows that $|b\rangle$ is an eigenket of H with eigenvalue $E_n + \hbar\omega$. The length-square of $|b\rangle$ is

$$\big|A^\dagger|E_n\rangle\big|^2 = \langle E_n|AA^\dagger|E_n\rangle = \langle E_n|\left(\frac{H}{\hbar\omega} + \tfrac{1}{2}\right)|E_n\rangle = \frac{E_n}{\hbar\omega} + \tfrac{1}{2}. \tag{3.8}$$

Now squeezing H between $\langle E_n|$ and $|E_n\rangle$ we find with (3.1) that

$$E_n = \langle E_n|H|E_n\rangle = \frac{1}{2m\omega}\left(\big|p|E_n\rangle\big|^2 + m^2\omega^2\big|x|E_n\rangle\big|^2\right) \geq 0. \tag{3.9}$$

Thus the energy eigenvalues are non-negative, so $\big|A^\dagger|E_n\rangle\big|^2 > 0$ and by repeated application of $A^\dagger$ we can construct an infinite series of eigenstates with energy $E_n + k\hbar\omega$ for $k = 0, 1, \ldots$

Similarly, we can show that provided $A|E_n\rangle$ has non-zero length-squared, it is an eigenket of H for energy $E_n - \hbar\omega$. Since we know that all eigenvalues are non-negative, for some energy E_0, $A|E_0\rangle$ must vanish. Equating to zero the length-squared of this vector we obtain an equation for E_0:

$$0 = \big|A|E_0\rangle\big|^2 = \langle E_0| \left(\frac{H}{\hbar\omega} - \tfrac{1}{2}\right) |E_0\rangle = \frac{E_0}{\hbar\omega} - \tfrac{1}{2}. \tag{3.10}$$

So $E_0 = \frac{1}{2}\hbar\omega$ and we have established that the eigenvalues of H are

$$\hbar\omega \times (\tfrac{1}{2}, \tfrac{3}{2}, \ldots, \tfrac{2r+1}{2}, \ldots) \quad \text{that is} \quad E_r = (r + \tfrac{1}{2})\hbar\omega. \tag{3.11}$$

The operators $A^\dagger$ and A with which we have obtained these important results are respectively called **creation** and annihilation operators because the first creates an excitation of the oscillator, and the second destroys one. In quantum field theory particles are interpreted as excitations of the vacuum and each particle species is associated with creation and annihilation operators that create and destroy particles of the given species. A and $A^\dagger$ are also called **ladder operators**.

We now examine the eigenkets of H. Let $|r\rangle$ denote the state of energy $(r + \frac{1}{2})\hbar\omega$. In this notation the lowest-energy state, or **ground state**, is $|0\rangle$ and its defining equation is $A|0\rangle = 0$. From equation (3.2a) this equation reads

$$0 = A|0\rangle = \frac{m\omega x|0\rangle + \mathrm{i}p|0\rangle}{\sqrt{2m\hbar\omega}}. \tag{3.12}$$

We now go to the position representation by multiplying through by $\langle x|$. With equations (2.48) and (2.49) we find that the equation becomes

$$\frac{1}{\sqrt{2m\hbar\omega}} \left(m\omega x + \hbar\frac{\partial}{\partial x}\right)\langle x|0\rangle = 0. \tag{3.13}$$

This is a linear, first-order differential equation. Its integrating factor is $\exp(m\omega x^2/2\hbar)$, so the correctly normalised wavefunction is

$$\langle x|0\rangle = \frac{1}{(2\pi\ell^2)^{1/4}}\, \mathrm{e}^{-x^2/4\ell^2}, \quad \text{where} \quad \ell \equiv \sqrt{\frac{\hbar}{2m\omega}}. \tag{3.14}$$

Notice that this solution is unique, so the ground state is **non-degenerate**. It is a Gaussian function, so the probability distribution $P(x) = |\langle x|0\rangle|^2$ for the position of the particle that forms the oscillator is also a Gaussian: its dispersion is ℓ.

From equations (2.63) and (2.64) we see that the momentum distribution of the wavefunction (3.14) is

$$P(p) \equiv |\langle p|0\rangle|^2 \propto \mathrm{e}^{-2\ell^2 p^2/\hbar^2}, \tag{3.15}$$

which is a Gaussian with dispersion $\sigma_p = \hbar/2\ell$. By inserting $x = \ell$ and $p = \sigma_p$ in the Hamiltonian (3.1) we obtain estimates of the typical kinetic and potential energies of the particle when it's in its ground state. We find that both energies are $\sim \frac{1}{4}\hbar\omega$. In fact one can straightforwardly show that $H(\ell, \sigma_p)$ is minimised subject to the constraint $\ell\sigma_p \geq \hbar/2$ when ℓ and σ_p take the values that we have derived for the ground state (Problem 3.4). In other words, in its ground state the particle is as stationary and as close to the origin as the uncertainty principle permits; there is a conflict between the advantage energetically of being near the origin, and the energetic penalty that the uncertainty principle exacts for having a well defined position.

Every system that has a confining potential exhibits an analogous **zero-point motion**. The energy tied up in this motion is called **zero-point energy**. Zero-point motion is probably the single most important prediction of quantum mechanics, for the material world is at every level profoundly influenced by this phenomenon.

We obtain the wavefunctions of excited states by applying powers of the differential operator $A^\dagger$ to $\langle x|0\rangle$. Equation (3.8) enables us to find the normalisation constant α in the equation $|n+1\rangle = \alpha A^\dagger|n\rangle$; it implies that $\alpha^2 = n+1$. the generalisation of equation (3.10) enables us to determine the number β in the equation $|n-1\rangle = \beta A|n\rangle$, and we have finally

$$|n+1\rangle = \frac{1}{\sqrt{n+1}}A^\dagger|n\rangle \quad ; \quad |n-1\rangle = \frac{1}{\sqrt{n}}A|n\rangle. \tag{3.16}$$

It is useful to remember that the normalisation constant is always the square root of the largest value of n appearing in the equation. As a specific example

$$\langle x|1\rangle = \frac{1}{\sqrt{2m\hbar\omega}}\left(m\omega x - \hbar\frac{\partial}{\partial x}\right)\langle x|0\rangle = \frac{1}{(2\pi\ell)^{1/4}}\frac{x}{\ell}\,\mathrm{e}^{-x^2/4\ell^2}. \tag{3.17}$$

Whereas the ground-state wavefunction is an even function of x, the wavefunction of the first excited state is an odd function because $A^\dagger$ is odd in x. Wavefunctions that are even in x are said to be of **even parity**, while those that are odd functions have **odd parity**. It is clear that this pattern will be repeated as we apply further powers of $A^\dagger$ to generate the other states of well-defined energy, so $\langle x|n\rangle$ is even parity if n is even, and odd parity otherwise.

Notice that the operator $N \equiv A^\dagger A$ is Hermitian. By equations (3.16) $N|n\rangle = n|n\rangle$, so its eigenvalue tells you the number of excitations the oscillator has. Hence N is called the **number operator**.

Let's use these results to find the mean-square displacement $\langle n|x^2|n\rangle$ when the oscillator is in its n^{th} excited state. Adding equations (3.2) we express x as a linear combination of A and $A^\dagger$

$$x = \sqrt{\frac{\hbar}{2m\omega}}(A + A^\dagger) = \ell(A + A^\dagger), \tag{3.18}$$

where ℓ is defined by (3.14), so

$$\langle n|x^2|n\rangle = \ell^2\langle n|(A + A^\dagger)^2|n\rangle. \tag{3.19}$$

When we multiply out the bracket on the right, the only terms that contribute are the ones that involve equal numbers of As and $A^\dagger$s. Thus

$$\langle n|x^2|n\rangle = \ell^2\langle n|(AA^\dagger + A^\dagger A)|n\rangle = \ell^2(2n+1) = \ell^2\frac{2E_n}{\hbar\omega}, \tag{3.20}$$

where we have used equations (3.16) and (3.11). If we use equation (3.14) to eliminate ℓ, we obtain a formula that is valid in classical mechanics (Problem 2.23).

3.2 Dynamics of oscillators

By equations (2.29) and (3.11), the $n^{\rm th}$ excited state of the harmonic oscillator evolves in time according to

$$|n,t\rangle = {\rm e}^{-{\rm i}(n+1/2)\omega t}|r,0\rangle \tag{3.21}$$

Consequently, no state oscillates at the oscillator's classical frequency ω. How do we reconcile this result with classical physics?

We have seen that we make the link from quantum to classical physics by considering the expectation values of observables – if classical physics applies, the measured value of any observable will lie close to the expectation value, so the latter provides an accurate description of what's happening. Equation (2.35) tells us that when a system is in an energy eigenstate, the expectation value of any time-independent observable Q cannot depend on time. Equation (3.21) enables us to obtain this result from a different point of view by showing that when we form the expectation value $\langle Q\rangle = \langle\psi|Q|\psi\rangle$, the factor ${\rm e}^{-{\rm i}E_nt/\hbar}$ in the ket $|\psi,t\rangle = {\rm e}^{-{\rm i}E_nt/\hbar}|E_n\rangle$ cancels on the corresponding factor in $\langle\psi,t|$. Hence energy eigenstates are incapable of motion.[2] The system is capable of motion only if there are non-negligible amplitudes to measure more than one possible energy, or, equivalently, if none of the a_i in the sum (2.32) has near unit modulus.

Consideration of the motion of a harmonic oscillator will make this general point clearer. If the oscillator's state is written

$$|\psi,t\rangle = \sum a_j\,{\rm e}^{-{\rm i}E_jt/\hbar}|j\rangle, \tag{3.22}$$

then the expectation value of x is

$$\langle x\rangle = \sum_{jk} a_k^*a_j{\rm e}^{{\rm i}(E_k-E_j)t/\hbar}\langle k|x|j\rangle = \sum_{jk} a_k^*a_j{\rm e}^{{\rm i}(k-j)\omega t}\langle k|x|j\rangle. \tag{3.23}$$

We simplify this expression by using equation (3.18) to replace x with $\ell(A+A^\dagger)$ and then using (3.16) to evaluate the matrix elements of A and $A^\dagger$:

$$\begin{aligned}\langle x\rangle &= \ell\sum_{jk} a_k^*a_j{\rm e}^{{\rm i}(k-j)\omega t}\langle k|(A+A^\dagger)|j\rangle\\ &= \ell\sum_{jk} a_k^*a_j{\rm e}^{{\rm i}(k-j)\omega t}\big(\sqrt{j}\langle k|j-1\rangle + \sqrt{j+1}\langle k|j+1\rangle.\end{aligned} \tag{3.24}$$

Since $\langle k|j-1\rangle$ vanishes unless $k=j-1$, it's now easy to perform the sum over k, leaving two terms to be summed over j. On account of the factor $\sqrt{j}$ we can restrict the first of these sums to $j>0$, and in the second sum we replace j by $j'\equiv j+1$ and then replace the symbol j' by j so we can combine the two sums. After these operations we have

$$\begin{aligned}\langle x\rangle &= \ell\sum_{j=1}\sqrt{j}(a_{j-1}^*a_j{\rm e}^{-{\rm i}\omega t} + a_j^*a_{j-1}{\rm e}^{{\rm i}\omega t})\\ &= \sum_j X_j\cos(\omega t+\phi_j),\end{aligned} \tag{3.25a}$$

where the real numbers X_j and ϕ_j are defined by

$$2\sqrt{j}\ell\, a_j^*a_{j-1} = X_j{\rm e}^{{\rm i}\phi_j}. \tag{3.25b}$$

---

[2] If we consider that t is the variable canonically conjugate to energy, this fact becomes a manifestation of the uncertainty principle.

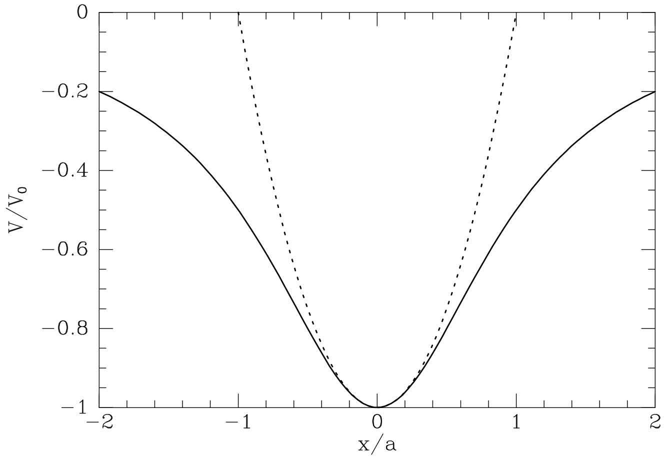


Figure 3.1 The potential energy $V(x)$ of an anharmonic oscillator (full curve) and $V(x)$ for the harmonic oscillator obtained by restricting the potential to the first two terms in its Maclaurin expansion (dashed curve).

Thus $\langle x\rangle$ oscillates sinusoidally at the classical frequency ω regardless of the amplitudes a_j. Thus we have recovered the classical result that the frequency at which a harmonic oscillator oscillates is independent of amplitude and equal to $\sqrt{k/m}$, where k is the oscillator's spring constant.

In the classical regime, the only non-negligible amplitudes a_j have indices j that cluster around some large number n. Consequently, a measurement of the energy is guaranteed to yield a value that lies close to $E = E_n$, and from equation (3.20) it follows that the mean value of x^2 will lie close to $\overline{x^2} = 2\ell^2 E_n/(\hbar\omega)$. Classically, the time average of x^2 is proportional to the average potential energy, which is just half the total energy. Hence, averaging the Hamiltonian (3.1) we conclude that classically $\overline{x^2} = E/(m\omega^2)$, in precise agrees with the quantum-mechanical result. The correspondence principle requires the classical and quantum-mechanical values of $\overline{x^2}$ to agree for large n. That they agree even for small n is a coincidence.

3.2.1 Anharmonic oscillators

The Taylor series of the potential energy $V(x)$ of a harmonic oscillator is very special: it contains precisely one non-trivial term, that proportional to x^2. Real oscillators have potential-energy functions that invariably deviate from this ideal to some degree. The deviation is generally in the sense that $V(x) < \frac{1}{2}V''(0)x^2$ for $x > 0$ – see Figure 3.1. One reason why deviations from harmonicity are generally of this type is that it takes only a finite amount of energy to break a real object, so $V(\infty)$ should be finite, whereas the potential energy function of a harmonic oscillator increases without limit as $x \to \infty$.

Consider the anharmonic oscillator that has potential energy

$$V(x) = -\frac{a^2 V_0}{a^2 + x^2}, \tag{3.26}$$

where V_0 and a are constants. We cannot find the stationary states of this oscillator analytically any more than we can analytically solve its classical equations of motion.[3] But we can determine its quantum mechanics numerically,[4] and doing so will help to show which aspects of the results we

---

[3] Murphy's law is in action here: the dynamics of the pendulum is analytically intractable precisely because it is richer and more interesting than that of the harmonic oscillator.

[4] A good way to do this is to turn the TISE into a finite matrix equation and then to

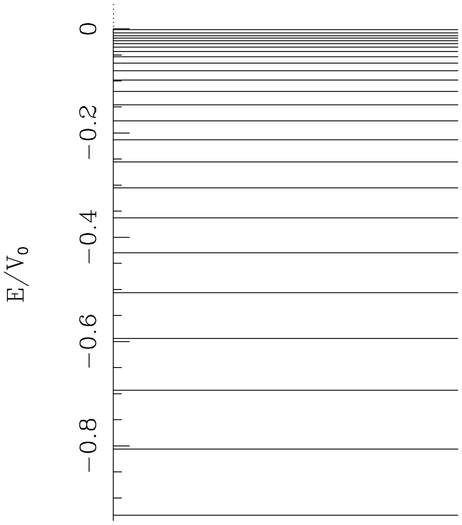


Figure 3.2 The spectrum of the anharmonic oscillator for which the potential is plotted in Figure 3.1 when the dimensionless variable $2ma^2V_0/\hbar^2 = 100$.

have obtained for the harmonic oscillator are special, and which have general applicability.

Figure 3.2 shows the anharmonic oscillator's energy spectrum. At low energies, when the pendulum is nearly harmonic, the energies are nearly uniformly spaced in E. As we proceed to higher energies, the spacing between levels diminishes, with the consequence that infinitely many energy levels are packed into the finite energy range between $-V_0$ and zero, where the particle becomes free. This crowding of the energy levels has the following implication for the time dependence of $\langle x\rangle$. Suppose there are just two energies with non-zero amplitudes, a_N and a_{N+1}. Then $\langle x\rangle$ will be given by

$$\langle x\rangle = a_N^* a_{N+1} \mathrm{e}^{\mathrm{i}(E_N - E_{N+1})t/\hbar} \langle N|x|N+1\rangle + \text{complex conjugate}. \tag{3.27}$$

This is a sinusoidal function of time, but its period, $T = h/(E_{N+1} - E_N)$, depends on N. If we increase the energy and amplitude of the oscillator, we will increase N and Figure 3.2 shows that T will also increase. Classically the period of the oscillator increases with amplitude in just the same way. Thus there is an intimate connection between the spacing of the energy levels and classical dynamics.

Consider now the case in which the energy is more uncertain, so that several of the a_j are non-zero, and let these non-zero a_j be clustered around $j = N$ (see Figure 3.3). In this case several terms will occur in the sum for $\langle x\rangle$

$$\begin{aligned}\langle x\rangle = \cdots &+ a_{N-1}^* a_N \mathrm{e}^{\mathrm{i}(E_{N-1} - E_N)t/\hbar} \langle N-1|x|N\rangle \\ &+ a_{N+1}^* a_N \mathrm{e}^{\mathrm{i}(E_{N+1} - E_N)t/\hbar} \langle N+1|x|N\rangle \\ &+ a_{N+3}^* a_N \mathrm{e}^{\mathrm{i}(E_{N+3} - E_N)t/\hbar} \langle N+3|x|N\rangle + \cdots\end{aligned} \tag{3.28}$$

where we have anticipated a result of §4.1.4 below that the matrix element $\langle j|x|k\rangle$ vanishes if $j - k$ is even. The sum (3.28) differs from the corresponding one (3.25a) for a harmonic oscillator in the presence of matrix elements $\langle j|x|k\rangle$ with $|j - k| > 1$: in the case of the harmonic oscillator these matrix elements vanish, but in the general case they won't. In consequence the series contains terms with frequencies $(E_{N+3} - E_N)/\hbar$ as well as terms

---

use a numerical linear-algebra package to find the eigenvalues of the matrix. Figure 3.2 was obtained using the approximation $\psi_n'' \simeq (\psi_{n+1} + \psi_{n-1} - 2\psi_n)/(2\Delta^2)$, where ψ_n denotes $\psi(n\Delta)$ with Δ a small increment in x. With this approximation the TISE becomes the eigenvalue equation of a tridiagonal matrix that has $\alpha^{-1}/\Delta^2 + V_n$ on the leading diagonal and $-\frac{1}{2}\alpha^{-2}/\Delta^2$ above and below this diagonal, where $\alpha = 2ma^2V_0/\hbar^2$.

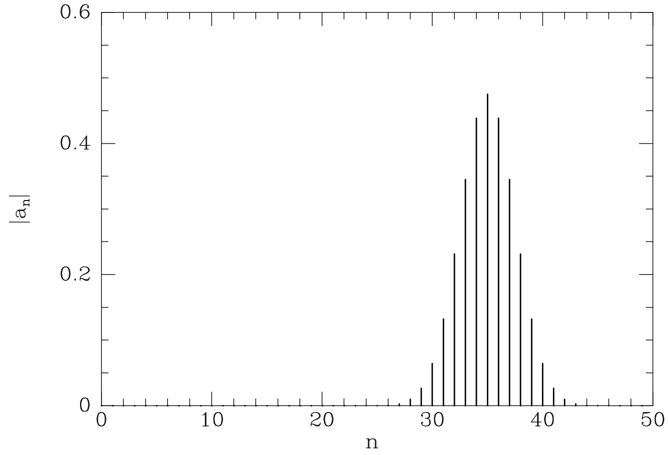


Figure 3.3 Values of a_j when there is significant uncertainty in E.

in $\omega_N \equiv (E_{N+1} - E_N)/\hbar$. If these additional frequencies were all integer multiples of a single frequency ω_N, the time dependence of $\langle x\rangle$ would be periodic with period $T_N = 2\pi/\omega_N$, but anharmonic, like that of the classical oscillator. Now $(E_{N+3} - E_N)/\hbar \simeq 3\omega_N$ because the spacing between energy levels changes only slowly with N, so when, as in Figure 3.3, the non-zero amplitudes are very tightly clustered around N, the additional frequencies will be integer multiples of ω_N to good accuracy, and the motion will indeed be periodic but anharmonic as classical mechanics predicts.

If we release the oscillator from near some large extension X, the non-negligible amplitudes a_j will be clustered around some integer N as depicted in Figure 3.3, and their phases will be such that at $t = 0$ the wavefunctions $\langle x|j\rangle$ will interfere constructively near X and sum to near zero elsewhere, ensuring that the mod-square of the wavefunction $\psi(x, 0) = \sum_j a_j\langle x|j\rangle$ is sharply peaked around $x = X$. At a general time the wavefunction will be given by

$$\psi(x,t) = \mathrm{e}^{-\mathrm{i}E_N t/\hbar}\sum_j \mathrm{e}^{\mathrm{i}(E_N - E_j)t/\hbar} a_j\langle x|j\rangle. \tag{3.29}$$

Since the spacing of the energy levels varies with index j, the frequencies in this sum will not be precisely equal to integer multiples of $\omega_N = (E_{N+1} - E_N)/\hbar$, so after an approximate period $T_N = 2\pi/\omega_N$ most terms in the series will not have quite returned to their values at $t = 0$. Consequently, the constructive interference around $x = X$ will be less sharply peaked than it was at $t = 0$, and the cancellation elsewhere will be correspondingly less complete. After each further approximate period T_N, the failure of terms in the series to return to their original values will be more marked, and the peak in $|\psi(x,t)|^2$ will be wider. After a long time $t \gg T_N$ the instants at which individual terms next return to their original values will be pretty uniformly distributed around an interval in t of length T_N, and $|\psi(x,t)|^2$ will cease to evolve very much: it will have become a smooth function throughout the range $|x| \lesssim X$.

This behaviour makes perfectly good sense classically. The uncertainty in E that enables the wavefunction to be highly localised at $t = 0$ corresponds in the classical picture to uncertainty in the initial displacement X. Since the period of an anharmonic oscillator is a function of the oscillator's energy, uncertainty in X implies uncertainty in the oscillator's period. After a long time even a small uncertainty in the period translates into a significant uncertainty in the oscillator's phase. Hence after a long time the probability distribution for the particle's position is fairly uniformly distributed within $|x| \leq X$ even in the classical case.

3.3 Motion in a magnetic field

The formalism we developed for a harmonic oscillator enables us to solve an important, and you might have thought unconnected, problem: the motion of a particle of mass m and charge Q in a uniform magnetic field of flux density B.

The first question to address when setting up the quantum-mechanical theory of a system is, "what's the Hamiltonian?" because it is the Hamiltonian that encodes mathematically how the system works, including what forces are acting. So we have to decide what the Hamiltonian should be for a particle of charge Q and mass m that moves in a magnetic field $\mathbf{B}(\mathbf{x})$. The answer proves to be

$$H = \frac{1}{2m}(\mathbf{p} - Q\mathbf{A})^2, \tag{3.30}$$

where $\mathbf{A}$ is the vector potential that generates $\mathbf{B}$ through $\mathbf{B} = \nabla \times \mathbf{A}$. The most persuasive theoretical motivation of this Hamiltonian involves relativity and lies beyond the scope of this book. However, since we are exploring a new and deeper level of physical theory, we can ultimately only proceed by making conjectures and then confronting the resulting predictions with experimental measurements. In this spirit we adopt equation (3.30) as a conjecture from which we can try to recover the known behaviour of a charged particle in a magnetic field. In subsequent chapters we will show that this formula accounts satisfactorily for features in the spectra of atoms. Hence we can be pretty sure that it is correct.

Since we know the equations of motion of a classical particle in a field $\mathbf{B}$, let's investigate the classical limit in the usual way, by finding the equations of motion of expectation values. With equation (2.34) we have that the rate of change of the expectation value of the i^{th} component of $\mathbf{x}$ is

$$\mathrm{i}\hbar \frac{\mathrm{d}\langle x_i\rangle}{\mathrm{d}t} = \langle [x_i, H]\rangle = \frac{1}{2m}\left\langle [x_i, (\mathbf{p} - Q\mathbf{A})^2]\right\rangle . \tag{3.31}$$

The rules (2.22) and the canonical commutation relation $[x_i, p_j] = \mathrm{i}\hbar\delta_{ij}$ enable us to simplify the commutator

$$\begin{aligned} 2m\mathrm{i}\hbar \frac{\mathrm{d}\langle x_i\rangle}{\mathrm{d}t} &= \langle [x_i, (\mathbf{p} - Q\mathbf{A})] \cdot (\mathbf{p} - Q\mathbf{A})\rangle \\ &\quad + \langle (\mathbf{p} - Q\mathbf{A}) \cdot [x_i, (\mathbf{p} - Q\mathbf{A})]\rangle \\ &= 2\mathrm{i}\hbar \langle p_i - QA_i\rangle , \end{aligned} \tag{3.32}$$

where we have used the fact that $\mathbf{x}$ commutes with $\mathbf{A}$ because $\mathbf{A}$ is a function of $\mathbf{x}$ only. Thus, with this Hamiltonian

$$\langle \mathbf{p}\rangle = m\frac{\mathrm{d}\langle \mathbf{x}\rangle}{\mathrm{d}t} + Q\langle \mathbf{A}\rangle . \tag{3.33}$$

that is, the momentum is $m\dot{\mathbf{x}}$ plus an amount proportional to the vector potential. It is possible to show that the additional term represents the momentum of the magnetic field that arises because the charge Q is moving (Problem 3.17).

In the classical limit we can neglect the difference between a variable and its expectation value because all uncertainties are small. Then with (3.33) the Hamiltonian (3.30) becomes just $\frac{1}{2}m\dot{\mathbf{x}}^2$, which makes perfect sense since we know that the Lorentz force $Q\dot{\mathbf{x}} \times \mathbf{B}$ does no work on a classical particle, and the particle's energy is just its kinetic energy.

To show that our proposed Hamiltonian generates the Lorentz force, we evaluate the rate of change of $\langle \mathbf{p}\rangle$:

$$\begin{aligned} \mathrm{i}\hbar \frac{\mathrm{d}\langle p_i\rangle}{\mathrm{d}t} = \langle [p_i, H]\rangle = \frac{1}{2m}\big\{ &\langle [p_i, (\mathbf{p} - Q\mathbf{A})] \cdot (\mathbf{p} - Q\mathbf{A})\rangle \\ &+ \langle (\mathbf{p} - Q\mathbf{A}) \cdot [p_i, (\mathbf{p} - Q\mathbf{A})]\rangle \big\}. \end{aligned} \tag{3.34}$$

We now use equation (2.25) to evaluate the commutator $[p_i, \mathbf{A}]$ and conclude that

$$\frac{\mathrm{d}\langle p_i\rangle}{\mathrm{d}t} = \frac{Q}{2m}\left\{\left\langle\frac{\partial\mathbf{A}}{\partial x_i}\cdot(\mathbf{p}-Q\mathbf{A})\right\rangle + \left\langle(\mathbf{p}-Q\mathbf{A})\cdot\frac{\partial\mathbf{A}}{\partial x_i}\right\rangle\right\}. \tag{3.35}$$

Notice that we cannot combine the two terms on the right of this equation because $\mathbf{p}$ does not commute with $\mathbf{A}$ and its derivatives. In the classical limit we can replace each operator by its expectation value, and then replace $\langle\mathbf{p}-Q\mathbf{A}\rangle$ by $m\,\langle\dot{\mathbf{x}}\rangle$. Similarly replacing the p on the left, we have in the classical limit

$$m\frac{\mathrm{d}^2x_i}{\mathrm{d}t^2} + Q\frac{\mathrm{d}A_i}{\mathrm{d}t} = Q\frac{\partial\mathbf{A}}{\partial x_i}\cdot\dot{\mathbf{x}}, \tag{3.36}$$

where we have omitted the expectation value signs that ought to be around every operator. The time derivative on the left is along the trajectory of the particle (i.e., to be evaluated at $\langle x\rangle_t$). If $\mathbf{A}$ has no explicit time dependence because the field $\mathbf{B}$ is static, its time derivative is just $\dot{\mathbf{x}}\cdot\nabla A_i$. We move this term to the right side and have

$$m\frac{\mathrm{d}^2x_i}{\mathrm{d}t^2} = Q\left(\dot{\mathbf{x}}\cdot\frac{\partial\mathbf{A}}{\partial x_i} - \dot{\mathbf{x}}\cdot\nabla A_i,\right) = Q\big\{\dot{\mathbf{x}}\times(\nabla\times\mathbf{A})\big\}_i. \tag{3.37}$$

Thus our proposed Hamiltonian (3.30) yields the Lorentz force in the classical limit.

3.3.1 Gauge transformations

Any magnetic field is **Gauge invariant**: $\mathbf{A}$ and $\mathbf{A}' = \mathbf{A} + \nabla\Lambda$ generate identical magnetic fields, where $\Lambda(\mathbf{x})$ is any scalar function. A potential problem with the Hamiltonian (3.30) is that it changes in a non-trivial way when we change gauge, which is worrying because H should embody the physics, which is independent of gauge. We now show that this behaviour gives rise to no physical difficulty providing we change the phases of all kets at the same time that we change the gauge in which we write $\mathbf{A}$. The idea that a change of gauge in a field such as $\mathbf{A}$ that mediates a force (in this case the electromagnetic force) requires a compensating change in the ket that is used to describe a given physical state, has enormously far-reaching ramifications in field theory.

Suppose $\psi(\mathbf{x}) = \langle\mathbf{x}|E\rangle$ is an eigenfunction of the Hamiltonian for $\mathbf{A}$:

$$(\mathbf{p}-Q\mathbf{A})^2|\psi\rangle = 2mE|\psi\rangle. \tag{3.38}$$

Then we show that

$$\phi(\mathbf{x}) \equiv \mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\psi(\mathbf{x}) \tag{3.39}$$

is an eigenfunction of the Hamiltonian we get by replacing $\mathbf{A}$ with $\mathbf{A}'$. We start by noting that

$$\mathbf{p} - Q\mathbf{A}' = \mathbf{p} - Q(\mathbf{A}+\nabla\Lambda) = (\mathbf{p} - Q\nabla\Lambda) - Q\mathbf{A}$$

and that for any wavefunction $\chi(\mathbf{x})$

$$\begin{aligned}\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\mathbf{p}\chi(\mathbf{x}) &= \mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\big(-\mathrm{i}\hbar\nabla\chi(\mathbf{x})\big)\\ &= -\mathrm{i}\hbar\nabla\big(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\chi\big) - Q\nabla\Lambda(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\chi)\\ &= (\mathbf{p}-Q\nabla\Lambda)\big(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\chi(\mathbf{x})\big).\end{aligned} \tag{3.40}$$

We subtract $Q\mathbf{A}\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\chi$ from each side to obtain

$$\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}(\mathbf{p}-Q\mathbf{A})\chi(\mathbf{x}) = (\mathbf{p}-Q\mathbf{A}-Q\nabla\Lambda)\big(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\chi(\mathbf{x})\big), \tag{3.41}$$

and then apply this result to $\chi \equiv (\mathbf{p} - Q\mathbf{A})\psi$:

$$\begin{aligned} \mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}(\mathbf{p} - Q\mathbf{A})^2\psi(\mathbf{x}) &= (\mathbf{p} - Q\mathbf{A} - Q\nabla\Lambda)\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}(\mathbf{p} - Q\mathbf{A})\psi(\mathbf{x}) \\ &= (\mathbf{p} - Q\mathbf{A} - Q\nabla\Lambda)^2\big(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\psi(\mathbf{x})\big), \end{aligned} \tag{3.42}$$

where the second equality uses (3.41) again, this time with χ put equal to ψ. So if

$$H\psi = \frac{1}{2m}(\mathbf{p} - Q\mathbf{A})^2\psi = E\psi, \tag{3.43}$$

then

$$H'\big(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\psi\big) = \frac{1}{2m}(\mathbf{p} - Q\mathbf{A} - Q\nabla\Lambda)^2\big(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\psi(\mathbf{x})\big) = E\big(\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\psi(\mathbf{x})\big), \tag{3.44}$$

In words, we can convert an eigenfunction of the Hamiltonian (3.30) with $\mathbf{A}$ to an eigenfunction of that Hamiltonian with $\mathbf{A}' = \mathbf{A} + \nabla\Lambda$ by multiplying it by $\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}$. Notice that Λ is an arbitrary function of $\mathbf{x}$, so multiplication by $\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}$ makes an entirely non-trivial change to $\psi(\mathbf{x})$.

Given that there is a one-to-one relation between the eigenfunctions of H before and after we make a gauge transformation, it is clear that the spectrum of energy levels must be unchanged by the gauge transformation. What about expectation values? Since both kets and the Hamiltonian undergo gauge transformations, we should be open to the possibility that other operators do too. Let R' be the gauge transform of the operator R. Then the expectation value of R is gauge invariant if

$$\langle R\rangle = \int \mathrm{d}^3\mathbf{x}\,\psi^*(\mathbf{x})R\psi(\mathbf{x}) = \int \mathrm{d}^3\mathbf{x}\,\psi^*(\mathbf{x})\mathrm{e}^{-\mathrm{i}Q\Lambda/\hbar}R'\mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\psi(\mathbf{x}). \tag{3.45}$$

Clearly this condition is satisfied for $R' = R$ if R is a function of $\mathbf{x}$ only. From our work above it is readily seen that if R depends on $\mathbf{p}$, the equation is satisfied if $\mathbf{p}$ only occurs through the combination $(\mathbf{p} - Q\mathbf{A})$, as in the Hamiltonian.[5] We believe that in any physical situation this condition on the occurrence of $\mathbf{p}$ will always be satisfied, so all expectation values are in fact gauge-invariant.

3.3.2 Landau Levels

We now find the stationary states of a spinless particle that moves in a uniform magnetic field. Let the z-axis be parallel to $\mathbf{B}$ and choose the gauge in which $\mathbf{A} = \frac{1}{2}B(-y, x, 0)$. Then from equation (3.30) we have

$$\begin{aligned} H &= \frac{1}{2m}\Big\{\big(p_x + \tfrac{1}{2}QBy\big)^2 + \big(p_y - \tfrac{1}{2}QBx\big)^2 + p_z^2\Big\} \\ &= \tfrac{1}{2}\hbar\omega(\pi_x^2 + \pi_y^2) + \frac{p_z^2}{2m}, \end{aligned} \tag{3.46a}$$

where $\omega = QB/m$ is the **Larmor frequency** and we have defined the dimensionless operators[6]

$$\pi_x \equiv \frac{p_x + \frac{1}{2}m\omega y}{\sqrt{m\omega\hbar}} \quad ; \quad \pi_y \equiv \frac{p_y - \frac{1}{2}m\omega x}{\sqrt{m\omega\hbar}}. \tag{3.46b}$$

H has broken into two parts. The term $p_z^2/2m$ is just the Hamiltonian of a free particle in one dimension – in §2.3.3 we already studied motion governed by this Hamiltonian. The part

$$H_{xy} \equiv \tfrac{1}{2}\hbar\omega(\pi_x^2 + \pi_y^2) \tag{3.46c}$$

---

[5] The principle that $\mathbf{p}$ and $\mathbf{A}$ only occur in the combination $\mathbf{p} - Q\mathbf{A}$ is known as the **principle of minimal coupling**.

[6] We are implicitly assuming that QB and therefore ω are positive. It is this assumption that leads to the angular momentum of a gyrating particle never being positive – see equation (3.57).

is essentially the Hamiltonian of a harmonic oscillator because it is the sum of squares of two Hermitian operators that satisfy the canonical commutation relation

$$[\pi_x,\pi_y] = \frac{1}{2\hbar}\big([y,p_y]-[p_x,x]\big) = \mathrm{i}. \tag{3.47}$$

The ladder operators are

$$\left.\begin{aligned} a &= \frac{1}{\sqrt{2}}(\pi_x+\mathrm{i}\pi_y) \\ a^\dagger &= \frac{1}{\sqrt{2}}(\pi_x-\mathrm{i}\pi_y) \end{aligned}\right. \quad\Rightarrow\quad [a,a^\dagger] = \mathrm{i}[\pi_y,\pi_x] = 1, \tag{3.48}$$

and in terms of them H_{xy} is

$$H_{xy} = \hbar\omega(a^\dagger a + \tfrac{1}{2}). \tag{3.49}$$

It follows that the energy levels are $E = \hbar\omega(\frac{1}{2},\frac{3}{2},\ldots)$. These discrete energy levels for a charged particle in a uniform magnetic field are known as **Landau levels**.

If particles can move freely parallel to $\mathbf{B}$ (which may not be possible in condensed-matter systems), the overall energy spectrum will be continuous notwithstanding the existence of discrete Landau levels.

In the case of an electron the Larmor frequency is usually called the **cyclotron frequency**. It evaluates to $176(B/1\,\mathrm{T})\,\mathrm{GHz}$, so the spacing of the energy levels is $1.16\times10^{-4}(B/1\,\mathrm{T})\,\mathrm{eV}$. At room temperature electrons have thermal energies of order $0.03\,\mathrm{eV}$, so the discreteness of Landau levels is usually experimentally significant in the laboratory only if the system is cooled to low temperatures and immersed in a strong magnetic field. The strongest magnetic fields known occur near neutron stars, where $B\sim10^8\,\mathrm{T}$ is not uncommon, and in these systems electrons moving from one Landau level to the next emit or absorb hard X-ray photons.

To find the wavefunction of a given Landau level, we write the ground state's defining equation in the position representation

$$a|0\rangle = 0 \quad\leftrightarrow\quad \left\{\hbar\Big(\frac{\partial}{\partial x}+\mathrm{i}\frac{\partial}{\partial y}\Big)+\tfrac{1}{2}m\omega(x+\mathrm{i}y)\right\}\langle\mathbf{x}|0\rangle = 0. \tag{3.50}$$

We transform to new coordinates $u\equiv x+\mathrm{i}y$, $v\equiv x-\mathrm{i}y$. The chain rule yields[7]

$$\frac{\partial}{\partial u} = \tfrac{1}{2}\Big(\frac{\partial}{\partial x}-\mathrm{i}\frac{\partial}{\partial y}\Big) \quad ; \quad \frac{\partial}{\partial v} = \tfrac{1}{2}\Big(\frac{\partial}{\partial x}+\mathrm{i}\frac{\partial}{\partial y}\Big), \tag{3.51}$$

so a and $a^\dagger$ can be written

$$a = -\mathrm{i}\frac{r_B}{\sqrt{2}}\Big(\frac{\partial}{\partial v}+\frac{u}{4r_B^2}\Big) \quad ; \quad a^\dagger = -\mathrm{i}\frac{r_B}{\sqrt{2}}\Big(\frac{\partial}{\partial u}-\frac{v}{4r_B^2}\Big), \tag{3.52a}$$

where

$$r_B \equiv \sqrt{\frac{\hbar}{m\omega}} = \sqrt{\frac{\hbar}{QB}}. \tag{3.52b}$$

Equation (3.50) now becomes

$$\frac{\partial\langle\mathbf{x}|0\rangle}{\partial v}+\frac{1}{4r_B^2}u\langle\mathbf{x}|0\rangle = 0. \tag{3.53}$$

---

[7] Aficianados of functions of a complex variable may ask what $\partial/\partial u$ can mean since the partial derivative involves holding constant v, which appears to be the complex conjugate of u. Use of u, v as independent coordinates requires permitting x, y to take on complex values. If you are nervous of using this mathematical fiction to solve differential equations, you should check that the wavefunction of equation (3.57) really is an eigenfunction of H.

Solving this first-order linear o.d.e. we find

$$\langle \mathbf{x}|0\rangle = g(u)\,\mathrm{e}^{-uv/4r_B^2} \tag{3.54}$$

where $g(u)$ is an arbitrary function. On account of the arbitrariness of $g(u)$, the ground state of motion in a magnetic field is not unique. This situation contrasts with the one we encountered when solving for the ground state of a harmonic oscillator. We obtain the simplest ground state by taking g to be a suitable normalising constant C – we'll consider more elaborate choices below. Our present ground-state wavefunction is

$$\langle \mathbf{x}|0\rangle = C\,\mathrm{e}^{-(x^2+y^2)/4r_B^2}. \tag{3.55}$$

In classical physics a particle that moves at speed v perpendicular to a uniform magnetic field moves in circles of radius $r = mv/QB = \sqrt{2mE}/QB = \sqrt{2E/m\omega^2}$. When $E = \frac{1}{2}\hbar\omega$ this radius agrees with the dispersion r_B in radius of the Gaussian probability distribution $|\langle x|0\rangle|^2$ that we have just derived.

A wavefunction in the first excited level is

$$\langle \mathbf{x}|1\rangle \propto \langle \mathbf{x}|a^\dagger|0\rangle \propto \left(\frac{\partial}{\partial u} - \frac{v}{4r_B^2}\right)\mathrm{e}^{-uv/4r_B^2} \propto v\mathrm{e}^{-uv/4r_B^2}. \tag{3.56}$$

It is easy to see that each further application of $a^\dagger$ will introduce an additional power of v, so that we have

$$\langle \mathbf{x}|n\rangle \propto v^n\mathrm{e}^{-uv/4r_B^2} = (x - \mathrm{i}y)^n\mathrm{e}^{-(x^2+y^2)/4r_B^2}. \tag{3.57}$$

We shall see in §7.2.3 that the factor $(x-\mathrm{i}y)^n$ implies that the particle has n units of angular momentum about the origin.[8] We can also show from this formula that for large n the expectation of the orbital radius increases as the square root of the energy, in agreement with classical mechanics (Problem 3.18).

Displacement of the gyrocentre A particle in the state (3.57) gyrates around the origin of the xy plane. Since the underlying physics (unlike the Hamiltonian 3.30) is invariant under displacements within the xy plane, there must be a ground-state ket in which the particle gyrates around any given point. Hence, every energy level associated with motion in a uniform magnetic field is highly **degenerate**: it has more than one linearly independent eigenket.

It was our choice of magnetic vector potential $\mathbf{A}$ that made the origin have a special status in H: the potential we used can be written $\mathbf{A} = \frac{1}{2}\mathbf{B}\times\mathbf{x}$. The choice $\Lambda = -\frac{1}{2}\mathbf{x}\cdot(\mathbf{B}\times\mathbf{a})$, where $\mathbf{a}$ is any vector, makes the gauge transformation from $\mathbf{A}$ to $\mathbf{A}' = \mathbf{A} - \frac{1}{2}\mathbf{B}\times\mathbf{a}$, so if $\mathbf{A} = \frac{1}{2}\mathbf{B}\times\mathbf{x}$, then $\mathbf{A}' = \frac{1}{2}\mathbf{B}\times(\mathbf{x}-\mathbf{a})$. If we replace $\mathbf{A}$ in H with $\mathbf{A}'$, it will prove expedient to redefine π_x, π_y such that the wavefunctions that are generated by the procedure we used before will describe a particle that gyrates around $\mathbf{x} = \mathbf{a}$ instead of the origin. Thus in the gauge $\mathbf{A}'$, the wavefunction of a ground-state particle that gyrates around $\mathbf{x} = \mathbf{a}$ is

$$\langle \mathbf{x}|0', \mathbf{a}\rangle = C\,\mathrm{e}^{-(\mathbf{x}-\mathbf{a})^2/4r_B^2}. \tag{3.58}$$

We can use the theory of gauge transformations that we derived in §3.3.1 to transform this back to our original gauge $\mathbf{A}$. The result is

$$\langle \mathbf{x}|0, \mathbf{a}\rangle = C\,\mathrm{e}^{\mathrm{i}Q(\mathbf{B}\times\mathbf{a})\cdot\mathbf{x}/2\hbar}\,\mathrm{e}^{-(\mathbf{x}-\mathbf{a})^2/4r_B^2}. \tag{3.59}$$

---

[8] This statement follows because in spherical polar coordinates $(x - \mathrm{i}y)^n = r^n\mathrm{e}^{-\mathrm{i}n\phi}$.

This procedure is easily generalised to the determination of the wavefunction of the $n^{\rm th}$ Landau level for gyration about $\mathbf{x} = \mathbf{a}$.

A complete set of mutually orthogonal stationary states is needed if we want to expand a general state of motion in a magnetic field as a linear combination of stationary states. Wavefunctions such as (3.55) and (3.59) that differ only in their gyrocentres are not orthogonal, so it is not convenient to combine them in a set of basis states. To obtain a complete set of mutually orthogonal states we can either return to equation (3.54) and set $g(u) = u,\ u^2, \ldots$, etc., or we can step still further back to equations (3.46) and note that we started with four operators, x, y, p_x and p_y, but expressed the Hamiltonian H_{xy} in terms of just two operators π_x and π_y, which we then packaged into the ladder operators a and $a^\dagger$.

Consider the operators

$$\xi_x \equiv \frac{p_x - \frac{1}{2}m\omega y}{\sqrt{m\omega\hbar}} \quad ; \quad \xi_y \equiv \frac{p_y + \frac{1}{2}m\omega x}{\sqrt{m\omega\hbar}}. \tag{3.60}$$

They differ from the operators π_x and π_y defined by equations (3.46b) only in a sign each, and they commute with them. For example

$$\begin{aligned}[][\xi_x, \pi_x] &= \frac{1}{m\omega\hbar}[p_x - \tfrac{1}{2}m\omega y, p_x + \tfrac{1}{2}m\omega y] = 0\\ [\xi_x, \pi_y] &= \frac{1}{m\omega\hbar}[p_x - \tfrac{1}{2}m\omega y, p_y - \tfrac{1}{2}m\omega x] = 0.\end{aligned} \tag{3.61}$$

Consequently they commute with H_{xy}. On the other hand, $[\xi_x, \xi_y] = -\mathrm{i}$, so from these operators we can construct the ladder operators

$$\left.\begin{aligned} b &= \frac{1}{\sqrt{2}}(\xi_x - \mathrm{i}\xi_y)\\ b^\dagger &= \frac{1}{\sqrt{2}}(\xi_x + \mathrm{i}\xi_y)\end{aligned}\right. \quad \Rightarrow \quad [b, b^\dagger] = \mathrm{i}[\xi_x, \xi_y] = 1. \tag{3.62}$$

Since these ladder operators commute with H_{xy}, we can find a complete set of mutual eigenkets of $b^\dagger b$ and H_{xy}.

In the position representation the new ladder operators are

$$b = -\mathrm{i}\frac{r_B}{\sqrt{2}}\Big(\frac{\partial}{\partial u} + \frac{v}{4r_B^2}\Big) \quad ; \quad b^\dagger = -\mathrm{i}\frac{r_B}{\sqrt{2}}\Big(\frac{\partial}{\partial v} - \frac{u}{4r_B^2}\Big). \tag{3.63}$$

When we apply b to the ground-state wavefunction $\langle\mathbf{x}|0\rangle = C\mathrm{e}^{-uv/4r_B^2}$ (eq. 3.55), we find

$$b\langle\mathbf{x}|0\rangle = -\mathrm{i}\frac{Cr_B}{\sqrt{2}}\Big(\frac{\partial}{\partial u} + \frac{v}{4r_B^2}\Big)\mathrm{e}^{-uv/4r_B^2} = 0. \tag{3.64}$$

Thus $C\mathrm{e}^{-uv/4r_B^2}$ is annihilated by both a and b. When we apply $b^\dagger$ to this wavefunction we obtain

$$b^\dagger\langle\mathbf{x}|0\rangle = -\mathrm{i}\frac{Cr_B}{\sqrt{2}}\Big(\frac{\partial}{\partial v} - \frac{u}{4r_B^2}\Big)\mathrm{e}^{-uv/4r_B^2} \propto u\,\mathrm{e}^{-uv/4r_B^2}, \tag{3.65}$$

which is the wavefunction we would have obtained if we had set $g(u) = u$ in equation (3.54). In fact it's clear from equation (3.65) that every application of $b^\dagger$ will introduce an additional factor u before the exponential. Therefore the series of ground-state wavefunctions that are obtained by repeatedly applying $b^\dagger$ to $\mathrm{e}^{-uv/4r_B^2}$ are all of the form

$$(b^\dagger)^n\langle\mathbf{x}|0\rangle \propto u^n\mathrm{e}^{-uv/4r_B^2}. \tag{3.66}$$

The only difference between this general ground-state wavefunction[9] and the wavefunction of the $n^{\rm th}$ excited Landau level (eq. 3.56) is that the former has u^n rather than v^n in front of the exponential. For the physical explanation of this result, see Problem 3.21.

---

[9] The absolute value of the real part of this is shown for $n = 4$ on the front cover.

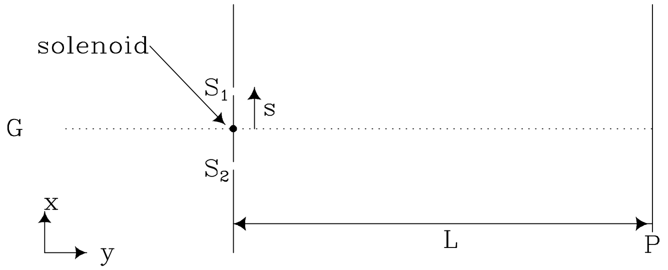


Figure 3.4 The Aharonov–Bohm experiment

3.3.3 Aharonov-Bohm effect

Imagine a very long, thin solenoid that runs parallel to the z axis. There is a strong magnetic field inside the solenoid, but because the solenoid is long, the field lines are extremely thinly spread as they return from the solenoid's north pole to its south pole, and outside the solenoid the magnetic field is negligibly small. In the limit that the solenoid becomes infinitely thin, a suitable vector potential for the field is

$$\mathbf{A} = \frac{\Phi}{2\pi}\left(-\frac{y}{r^2}, \frac{x}{r^2}, 0\right), \tag{3.67}$$

where $r = \sqrt{x^2+y^2}$ and Φ is the magnetic flux through the solenoid. To justify this statement we note that when we integrate $\mathbf{A}$ around a circle of radius r, the integral evaluates to Φ independent of r. But by Stokes' theorem

$$\oint \mathrm{d}\mathbf{x}\cdot\mathbf{A} = \int \mathrm{d}^2\mathbf{x}\cdot\nabla\times\mathbf{A} = \int \mathrm{d}^2\mathbf{x}\cdot\mathbf{B}. \tag{3.68}$$

Thus Φ units of flux run along the axis $r = 0$, and there is no flux anywhere else.

Now we place a screen with two slits in the plane $y = 0$, with the slits distance $2s$ apart and running parallel to the solenoid and on either side of it. We bombard the screen from $y < 0$ with particles that have well defined momentum $\mathbf{p} = p\mathbf{j}$ parallel to the y axis, and we detect the arrival of the particles on a screen P that lies in the plane $y = L$ – apart from the presence of the solenoid, the arrangement is identical to that of the standard two-slit experiment of §2.3.4. Classical physics predicts that the particles are unaffected by $\mathbf{B}$ since they never enter the region of non-zero $\mathbf{B}$. Aharonov & Bohm pointed out[10] that the prediction of quantum mechanics is different.

Consider the function

$$\Lambda = -\frac{\Phi}{2\pi}\theta, \tag{3.69}$$

where θ is the usual polar angle in the xy plane. Since $\theta = \arctan(y/x)$,

$$\frac{\partial\theta}{\partial x} = -\frac{y}{r^2} \quad\text{and}\quad \frac{\partial\theta}{\partial y} = \frac{x}{r^2}, \tag{3.70}$$

and the gradient of Λ is

$$\nabla\Lambda = -\frac{\Phi}{2\pi r^2}(-y, x, 0) \tag{3.71}$$

which is minus the vector potential $\mathbf{A}$ of equation (3.67). So let's make a gauge transformation from $\mathbf{A}$ to $\mathbf{A}' = \mathbf{A} + \nabla\Lambda$. In this gauge, the vector potential vanishes, so the Hamiltonian is just that of a free particle, $\mathbf{p}^2/2m$. Hence the analysis of §2.3.4 applies, and the amplitude to pass through a

---

[10] Y. Aharonov & D. Bohm, *Phys. Rev.* **115**, 485 (1959)

given slit and arrive at a point on the screen P with coordinate x has a phase ϕ that is proportional to x (cf eq. 2.71):

$$\phi_i = \text{constant} \pm \frac{psx}{\hbar L}, \tag{3.72}$$

where the plus sign applies for one slit and the minus sign for the other.

Our choice of gauge leads to a tricky detail, however. We require Λ to be single valued, so we must restrict the polar angle θ to a range 2π in extent. Consequently, θ and Λ must be somewhere discontinuous. We get around this problem by using different forms of Λ, and therefore different gauges, to derive the amplitudes for arrival at x from each slit. For slit S_1 at $x = +s$, we take $-\pi < \theta \le \pi$, and for S_2 at $x = -s$ we take $-2\pi < \theta \le 0$. With these choices the discontinuity in Λ occurs where the electron does *not* go, and Λ is always the same in the region $y < 0$ occupied by the incoming electron beam. Consequently, the amplitudes for arrival at a point x on the screen P are the same as if the solenoid were not there. However, before we can add the amplitudes and calculate the interference pattern, we have to transform to a common gauge. The easiest way to do this is to transform the amplitude for S_2 to the gauge of S_1. The function that effects the transformation between the gauges is $\Delta \equiv \Lambda_1 - \Lambda_2$, where Λ_i is the gauge function used for slit S_i. At any point of P the two forms of θ differ by 2π, so $\Delta = -\Phi$. Therefore equation (3.39) requires us to multiplying the amplitude for S_2 by $\exp(-\mathrm{i}Q\Phi/\hbar)$, and the quantum interference term (1.15) becomes

$$\text{constant} \times \mathrm{e}^{\mathrm{i}(\phi_1 - \phi_2 + Q\Phi/\hbar)} \propto \exp\left(\frac{\mathrm{i}}{\hbar}\left\{\frac{2psx}{L} + Q\Phi\right\}\right). \tag{3.73}$$

The term $Q\Phi$ in the exponential shifts the centre of the interference pattern by an amount $\Delta x = -LQ\Phi/2ps$, so by switching the current in the solenoid on and off you can change the interference pattern that is generated by particles that never enter the region to which $\mathbf{B}$ is confined. This prediction was first confirmed experimentally by R.G. Chambers.[11] Although this effect has no counterpart in classical mechanics, curiously the shift Δx is independent of h and does not vanish in the limit $h \to 0$, which is often regarded as the classical limit.

Problems

3.1 After choosing units in which everything, including $\hbar = 1$, the Hamiltonian of a harmonic oscillator may be written $H = \frac{1}{2}(p^2 + x^2)$, where $[x, p] = \mathrm{i}$. Show that if $|\psi\rangle$ is a ket that satisfies $H|\psi\rangle = E|\psi\rangle$, then

$$\tfrac{1}{2}(p^2 + x^2)(x \mp \mathrm{i}p)|\psi\rangle = (E \pm 1)(x \mp \mathrm{i}p)|\psi\rangle. \tag{3.74}$$

Explain how this algebra enables one to determine the energy eigenvalues of a harmonic oscillator.

3.2 Given that $A|E_n\rangle = \alpha|E_{n-1}\rangle$ and $E_n = (n + \frac{1}{2})\hbar\omega$, where the annihilation operator of the harmonic oscillator is

$$A \equiv \frac{m\omega x + \mathrm{i}p}{\sqrt{2m\hbar\omega}}, \tag{3.75}$$

show that $\alpha = \sqrt{n}$. Hint: consider $|A|E_n\rangle|^2$.

3.3 The pendulum of a grandfather clock has a period of 1 s and makes excursions of 3 cm either side of dead centre. Given that the bob weighs 0.2 kg, around what value of n would you expect its non-negligible quantum amplitudes to cluster?

---

[11] *Phys. Rev. Lett.* **5**, 3 (1960)

3.4 Show that the minimum value of $E(p,x) \equiv p^2/2m + \frac{1}{2}m\omega^2x^2$ with respect to the real numbers p, x when they are constrained to satisfy $xp = \frac{1}{2}\hbar$, is $\frac{1}{2}\hbar\omega$. Explain the physical significance of this result.

3.5 How many nodes are there in the wavefunction $\langle x|n\rangle$ of the n^{th} excited state of a harmonic oscillator?

3.6 Show for a harmonic oscillator that the wavefunction of the second excited state is $\langle x|2\rangle = \text{constant} \times (x^2/\ell^2 - 1)\mathrm{e}^{-x^2/4\ell^2}$, where $\ell \equiv \sqrt{\frac{\hbar}{2m\omega}}$, and find the normalising constant. Hint: apply $A^\dagger$ to $|0\rangle$ twice in the position representation.

3.7 Use

$$x = \sqrt{\frac{\hbar}{2m\omega}}(A + A^\dagger) = \ell(A + A^\dagger) \tag{3.76}$$

to show for a harmonic oscillator that in the energy representation the operator x is

$$x_{jk} = \ell \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \cdots & & & & \\ \sqrt{1} & 0 & \sqrt{2} & 0 & & & & & \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \cdots & & & & \\ & & \sqrt{3} & \cdots & & & & & \\ \cdots & & \cdots & & \cdots & & \cdots & & \\ & & & \cdots & 0 & \sqrt{n-1} & \cdots & & \\ & & & & \sqrt{n-1} & 0 & \sqrt{n} & & \\ & & & & & \sqrt{n} & 0 & \sqrt{n+1} & \cdots \\ & & & & & & \sqrt{n+1} & 0 & \\ \cdots & & \cdots & & \cdots & & \cdots & & \cdots \end{pmatrix} \tag{3.77}$$

Calculate the same entries for the matrix p_{jk}.

3.8 At $t = 0$ the state of a harmonic oscillator of mass m and frequency ω is

$$|\psi\rangle = \tfrac{1}{2}|N-1\rangle + \tfrac{1}{\sqrt{2}}|N\rangle + \tfrac{1}{2}|N+1\rangle. \tag{3.78}$$

Calculate the expectation value of x as a function of time and interpret your result physically in as much detail as you can.

3.9* By expressing the annihilation operator A of the harmonic oscillator in the momentum representation, obtain $\langle p|0\rangle$. Check that your expression agrees with that obtained from the Fourier transform of

$$\langle x|0\rangle = \frac{1}{(2\pi\ell^2)^{1/4}}\,\mathrm{e}^{-x^2/4\ell^2}, \quad \text{where} \quad \ell \equiv \sqrt{\frac{\hbar}{2m\omega}}. \tag{3.79}$$

3.10 Show that for any two $N \times N$ matrices A, B, $\text{trace}([A,B]) = 0$. Comment on this result in the light of the results of Problem 3.7 and the canonical commutation relation $[x,p] = \mathrm{i}\hbar$.

3.11* A **Fermi oscillator** has Hamiltonian $H = f^\dagger f$, where f is an operator that satisfies

$$f^2 = 0 \qquad ; \qquad ff^\dagger + f^\dagger f = 1. \tag{3.80}$$

Show that $H^2 = H$, and thus find the eigenvalues of H. If the ket $|0\rangle$ satisfies $H|0\rangle = 0$ with $\langle 0|0\rangle = 1$, what are the kets (a) $|a\rangle \equiv f|0\rangle$, and (b) $|b\rangle \equiv f^\dagger|0\rangle$?

In quantum field theory the vacuum is pictured as an assembly of oscillators, one for each possible value of the momentum of each particle type. A boson is an excitation of a harmonic oscillator, while a fermion in an excitation of a Fermi oscillator. Explain the connection between the spectrum of $f^\dagger f$ and the Pauli principle.

3.12 In the time interval $(t+\delta t, t)$ the Hamiltonian H of some system varies in such a way that $|H|\psi\rangle|$ remains finite. Show that under these circumstances $|\psi\rangle$ is a continuous function of time.

A harmonic oscillator with frequency ω is in its ground state when the stiffness of the spring is instantaneously reduced by a factor $f^4 < 1$, so its natural frequency becomes $f^2\omega$. What is the probability that the oscillator is subsequently found to have energy $\frac{3}{2}\hbar f^2\omega$? Discuss the classical analogue of this problem.

3.13* P is the probability that at the end of the experiment described in Problem 3.12, the oscillator is in its second excited state. Show that when $f=\frac{1}{2}$, $P=0.144$ as follows. First show that the annihilation operator of the original oscillator

$$A = \tfrac{1}{2}\left\{(f^{-1}+f)A' + (f^{-1}-f)A'^{\dagger}\right\}, \tag{3.81}$$

where A' and $A'^{\dagger}$ are the annihilation and creation operators of the final oscillator. Then writing the ground-state ket of the original oscillator as a sum $|0\rangle = \sum_n c_n|n'\rangle$ over the energy eigenkets of the final oscillator, impose the condition $A|0\rangle = 0$. Finally use the normalisation of $|0\rangle$ and the orthogonality of the $|n'\rangle$. What value do you get for the probability of the oscillator remaining in the ground state?

Show that at the end of the experiment the expectation value of the energy is $0.2656\hbar\omega$. Explain physically why this is less than the original ground-state energy $\frac{1}{2}\hbar\omega$.

This example contains the physics behind the inflationary origin of the Universe: gravity explosively enlarges the vacuum, which is an infinite collection of harmonic oscillators (Problem 3.11). Excitations of these oscillators correspond to elementary particles. Before inflation the vacuum is unexcited so every oscillator is in its ground state. At the end of inflation, there is non-negligible probability of many oscillators being excited and each excitation implies the existence of a newly created particle.

3.14* In terms of the usual ladder operators A, $A^{\dagger}$, a Hamiltonian can be written

$$H = \mu A^{\dagger}A + \lambda(A + A^{\dagger}). \tag{3.82}$$

What restrictions on the values of the numbers μ and λ follow from the requirement for H to be Hermitian?

Show that for a suitably chosen operator B, H can be rewritten

$$H = \mu B^{\dagger}B + \text{constant}. \tag{3.83}$$

where $[B, B^{\dagger}] = 1$. Hence determine the spectrum of H.

3.15* Numerically calculate the spectrum of the anharmonic oscillator shown in Figure 3.2. From it estimate the period at a sequence of energies. Compare your quantum results with the equivalent classical results.

3.16* Let $B = cA + sA^{\dagger}$, where $c \equiv \cosh\theta$, $s \equiv \sinh\theta$ with θ a real constant and A, $A^{\dagger}$ are the usual ladder operators. Show that $[B, B^{\dagger}] = 1$.

Consider the Hamiltonian

$$H = \epsilon A^{\dagger}A + \tfrac{1}{2}\lambda(A^{\dagger}A^{\dagger} + AA), \tag{3.84}$$

where ϵ and λ are real and such that $\epsilon > \lambda > 0$. Show that when

$$\epsilon c - \lambda s = Ec \quad ; \quad \lambda c - \epsilon s = Es \tag{3.85}$$

with E a constant, $[B, H] = EB$. Hence determine the spectrum of H in terms of ϵ and λ.

3.17* *This problem is all classical emag, but it gives physical insight into quantum physics. It is hard to do without a command of Cartesian tensor notation.* A point charge Q is placed at the origin in the magnetic field generated by a spatially confined current distribution. Given that

$$\mathbf{E} = \frac{Q}{4\pi\epsilon_0}\frac{\mathbf{r}}{r^3} \tag{3.86}$$

and $\mathbf{B} = \nabla\times\mathbf{A}$ with $\nabla\cdot\mathbf{A} = 0$, show that the field's momentum

$$\mathbf{P} \equiv \epsilon_0 \int \mathrm{d}^3\mathbf{x}\,\mathbf{E}\times\mathbf{B} = Q\mathbf{A}(0). \tag{3.87}$$

Write down the relation between the particle's position and momentum and interpret this relation physically in light of the result you have just obtained.

Hint: write $\mathbf{E} = -(Q/4\pi\epsilon_0)\nabla r^{-1}$ and $\mathbf{B} = \nabla\times\mathbf{A}$, expand the vector triple product and integrate each of the resulting terms by parts so as to exploit in one $\nabla\cdot\mathbf{A} = 0$ and in the other $\nabla^2 r^{-1} = -4\pi\delta^3(\mathbf{r})$. The tensor form of Gauss's theorem states that $\int \mathrm{d}^3\mathbf{x}\,\nabla_i\mathbf{T} = \oint \mathrm{d}^2S_i\,\mathbf{T}$ no matter how many indices the tensor $\mathbf{T}$ may carry.

3.18* From equation (3.57) show that the the normalised wavefunction of a particle of mass m that is in the n^{th} Landau level of a uniform magnetic field B is

$$\langle\mathbf{x}|n\rangle = \frac{r^n \mathrm{e}^{-r^2/4r_B^2}\mathrm{e}^{-\mathrm{i}n\phi}}{2^{(n+1)/2}\sqrt{n!}\,\pi\, r_B^{n+1}}, \tag{3.88}$$

where $r_B = \sqrt{\hbar/QB}$. Hence show that the expectation of the particle's gyration radius is

$$\langle r\rangle_n \equiv \langle n|r|n\rangle = \sqrt{2}\{(n+\tfrac{1}{2})(n-\tfrac{1}{2})\times\cdots\times\tfrac{1}{2}\}\frac{r_B}{n!}. \tag{3.89}$$

Show further that

$$\frac{\delta \ln\langle r\rangle_n}{\delta n} \simeq \frac{1}{2n} \tag{3.90}$$

and thus show that in the limit of large n, $\langle r\rangle \propto \sqrt{E}$, where E is the energy of the level. Show that this result is in accordance with the correspondence principle.

3.19 A particle of charge Q is confined to move in the xy plane, with electrostatic potential $\phi = 0$ and vector potential $\mathbf{A}$ satisfying

$$\boldsymbol{\nabla}\times\mathbf{A} = (0,0,B). \tag{3.91}$$

Consider the operators ρ_x, ρ_y, R_x and R_y, defined by

$$\boldsymbol{\rho} = \frac{1}{QB}\,\hat{\mathbf{e}}_z\times(\mathbf{p}-Q\mathbf{A}) \qquad \text{and} \qquad \mathbf{R} = \mathbf{r}-\boldsymbol{\rho}, \tag{3.92}$$

where $\mathbf{r}$ and $\mathbf{p}$ are the usual position and momentum operators, and $\hat{\mathbf{e}}_z$ is the unit vector along $\mathbf{B}$. Show that the only non-zero commutators formed from the x- and y-components of these are

$$[\rho_x,\rho_y] = \mathrm{i}r_B^2 \qquad \text{and} \qquad [R_x,R_y] = -\mathrm{i}r_B^2, \tag{3.93}$$

where $r_B^2 = \hbar/QB$.

The operators a, $a^\dagger$, b and $b^\dagger$ are defined via

$$a = \frac{1}{\sqrt{2}\,r_B}(\rho_x+\mathrm{i}\rho_y) \qquad \text{and} \qquad b = \frac{1}{\sqrt{2}\,r_B}(R_y+\mathrm{i}R_x). \tag{3.94}$$

Evaluate $[a, a\dagger]$ and $[b, b^\dagger]$. Show that for suitably defined ω, the Hamiltonian can be written

$$H = \hbar\omega\left(a^\dagger a + \tfrac{1}{2}\right). \tag{3.95}$$

Given that there exists a unique state $|\psi\rangle$ satisfying

$$a|\psi\rangle = b|\psi\rangle = 0, \tag{3.96}$$

what conclusions can be drawn about the allowed energies of the Hamiltonian and their degeneracies? What is the physical interpretation of these results?

3.20* Equation (2.96) gives the probability current density associated with a wavefunction. Show that the flux given by this expression changes when we make a gauge change $\psi \to \psi' = \mathrm{e}^{\mathrm{i}Q\Lambda/\hbar}\psi$. Why is this state of affairs physically unacceptable?

Show that in Dirac's notation equation (2.94) is

$$\mathbf{J}(\mathbf{x}) = \frac{1}{m}\Re\left(\langle\psi|\mathbf{x}\rangle\langle\mathbf{x}|\mathbf{p}|\psi\rangle\right). \tag{3.97}$$

Modify this expression to obtain a gauge-invariant definition of **J**. Explain why your new expression makes good physical sense given the form that the kinetic-energy operator takes in the presence of a magnetic field. Show that in terms of the amplitude and phase θ of ψ your expression reads

$$\mathbf{J} = \frac{|\psi|^2}{m}(\hbar\boldsymbol{\nabla}\theta - Q\mathbf{A}). \tag{3.98}$$

Explicitly check that this formula makes **J** invariant under a gauge transformation.

Using cylindrical polar coordinates (r, ϕ, z), show that the probability current density associated with the wavefunction (3.88) of the n^{th} Landau level is

$$\mathbf{J}(r) = -\frac{\hbar r^{2n-1}\mathrm{e}^{-r^2/2r_B^2}}{2^{n+1}\pi n! m r_B^{2n+2}}\left(n + \frac{r^2}{2r_B^2}\right)\hat{\mathbf{e}}_\phi, \tag{3.99}$$

where $r_B \equiv \sqrt{\hbar/QB}$. Plot **J** as a function of r and interpret your plot physically.

3.21* Determine the probability current density associated with the n^{th} Landau ground-state wavefunction (3.66) (which for $n = 4$ is shown on the front cover). Use your result to explain in as much detail as you can why this state can be interpreted as a superposition of states in which the electron gyrates around different gyrocentres. Hint: adapt equation (3.99).

Why is the energy of a gyrating electron incremented if we multiply the wavefunction $\mathrm{e}^{-(m\omega/4\hbar)r^2}$ by $v^n = (x - \mathrm{i}y)^n$ but not if we multiply it by $u^n = (x + \mathrm{i}y)^n$?

3.22* In classical electromagnetism the magnetic moment of a planar loop of wire that has area A, normal $\hat{\mathbf{n}}$ and carries a current I is defined to be

$$\boldsymbol{\mu} = IA\hat{\mathbf{n}}. \tag{3.100}$$

Use this formula and equation (3.99) to show that the magnetic moment of a charge Q that is in a Landau level of a magnetic field B has magnitude $\mu = E/B$, where E is the energy of the level. Rederive this formula from classical mechanics.

4

Transformations & Observables

In §2.1 we associated an operator with every observable quantity through a sum over all states in which the system has a well-defined value of the observable (eq. 2.5). We found that this operator enabled us to calculate the expectation value of any function of the observable. Moreover, from the operator we could recover the observable's allowed values and the associated states because they are the operator's eigenvalues and eigenkets. These properties make an observable's operator a useful repository of information about the observable, a handy filing system. But they do not give the operator much physical meaning. Above all, they don't answer the question 'what does an operator actually *do* when it operates?' In this chapter we answer this question. In the process of doing this, we will see why the canonical commutation relations (2.54) have the form that they do, and introduce the angular-momentum operators, which will play important roles in the rest of the book.

4.1 Transforming kets

When one meets an unfamiliar object, one may study it by moving it around, perhaps turning it over in one's hands so as to learn about its shape. In §1.3.2 we claimed that all physical information about any system is encapsulated in its ket $|\psi\rangle$, so we must learn how $|\psi\rangle$ changes as we move and turn the system.

Even the simplest systems can have orientations in addition to positions. For example, an electron, a lithium nucleus or a water molecule all have orientations because they are not spherically symmetric: an electron is a magnetic dipole, a ^{7}Li nucleus has an electric quadrupole, and a water molecule is a V-shaped thing. The ket $|\psi\rangle$ that describes any of these objects contains information about the object's orientation in addition to its position and momentum. In the next subsection we shall focus on the location of a quantum system, but later we shall be concerned with its orientation as well, and in preparation for that work we explicitly display a label μ of the system's orientation and any other relevant properties, such as internal energy. For the moment μ is just an abstract symbol for orientation information; the details will be fleshed out in §7.1.

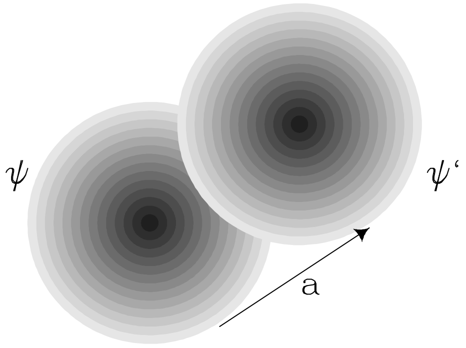


Figure 4.1 A spherical wavefunction and its displaced version.

4.1.1 Translating kets

We now focus on the location of our system. To keep track of this we use a coordinate system Σ_0 whose origin is some well-defined point, say the centre of our laboratory. We can investigate $|\psi\rangle$ by expanding it in terms of a complete set of eigenstates $|\mathbf{x},\mu\rangle$, where $\mathbf{x}$ is the position vector of the centre of mass and μ represents the system's orientation. The amplitude for finding the system's centre of mass at $\mathbf{x}$ with the orientation specified by μ is

$$\psi_\mu(\mathbf{x}) \equiv \langle \mathbf{x},\mu|\psi\rangle. \tag{4.1}$$

If we know all the derivatives of the wavefunction ψ_μ at a position $\mathbf{x}$, Taylor's theorem gives the value of the wavefunction at some other location $\mathbf{x}-\mathbf{a}$ as

$$\begin{aligned}\psi_\mu(\mathbf{x}-\mathbf{a}) &= \left[1-\mathbf{a}\cdot\frac{\partial}{\partial\mathbf{x}}+\frac{1}{2!}\left(\mathbf{a}\cdot\frac{\partial}{\partial\mathbf{x}}\right)^2-\dots\right]\psi_\mu(\mathbf{x})\\ &=\exp\left(-\mathbf{a}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\psi_\mu(\mathbf{x})\\ &=\langle\mathbf{x},\mu|\exp\left(-\mathrm{i}\frac{\mathbf{a}\cdot\mathbf{p}}{\hbar}\right)|\psi\rangle.\end{aligned} \tag{4.2}$$

This equation tells us that in the state $|\psi\rangle$, the amplitude to find the system at $\mathbf{x}-\mathbf{a}$ with orientation etc μ is the same as the amplitude to find the system with unchanged orientation at $\mathbf{x}$ when it is a *different* state, namely

$$|\psi'\rangle \equiv U(\mathbf{a})|\psi\rangle \quad \text{where} \quad U(\mathbf{a}) \equiv \exp(-\mathrm{i}\mathbf{a}\cdot\mathbf{p}/\hbar). \tag{4.3}$$

In this notation, equation (4.2) becomes

$$\psi_\mu(\mathbf{x}-\mathbf{a}) = \langle\mathbf{x},\mu|U(\mathbf{a})|\psi\rangle = \langle\mathbf{x},\mu|\psi'\rangle = \psi'_\mu(\mathbf{x}), \tag{4.4}$$

so, as Figure 4.1 illustrates, the wavefunction ψ'_μ for $|\psi'\rangle$ is the wavefunction we would expect for a system that is identical with the one described by $|\psi\rangle$ except for being shifted along the vector $\mathbf{a}$. We shall refer to this new system as the **translated** or transformed system and we shall say that the **translation operator** operator $U(\mathbf{a})$ translates $|\psi\rangle$ through $\mathbf{a}$ even though $|\psi\rangle$ is not an object in real space, so this is a slight abuse of language.

The ket $|\psi'\rangle$ of the translated system is a function of the vector $\mathbf{a}$. It is instructive to take its partial derivative with respect to one component of $\mathbf{a}$, say a_x. Evaluating the resulting derivative at $\mathbf{a}=0$, when $|\psi'\rangle = |\psi\rangle$, we find

$$\mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial a_x} = -\mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial x} = p_x|\psi\rangle. \tag{4.5}$$

Thus the operator p_x gives the rate at which the system's ket changes as we translate the system along the x axis. So we have answered the question

Box 4.1: Passive transformations

We can describe objects such as atoms equally well using any coordinate system. Imagine a whole family of coordinate systems set up throughout space, such that every physical point is at the origin of one coordinate system. We label by $\Sigma_{\mathbf{y}}$ the coordinate system whose origin coincides with the point labelled by $\mathbf{y}$ in our original coordinate system Σ_0, and we indicate the coordinate system used to obtain a wavefunction by making $\mathbf{y}$ a second argument of the wavefunction; $\psi_\mu(\mathbf{x};\mathbf{y})$ is the amplitude to find the system at the point labelled $\mathbf{x}$ in $\Sigma_{\mathbf{y}}$. Because the different coordinate systems vary smoothly with $\mathbf{y}$, we can use Taylor's theorem to express amplitudes in, say, $\Sigma_{\mathbf{a}+\mathbf{y}}$ in terms of amplitudes in $\Sigma_{\mathbf{y}}$. We have

$$\psi_\mu(\mathbf{x};\mathbf{a}+\mathbf{y}) = \exp\left(\mathbf{a}\cdot\frac{\partial}{\partial\mathbf{y}}\right)\psi_\mu(\mathbf{x};\mathbf{y}). \tag{1}$$

Now $\psi_\mu(\mathbf{x};\mathbf{a}) = \psi_\mu(\mathbf{x}+\mathbf{a};0)$ because both expressions give the amplitude for the system to be at the same physical location, the point called $\mathbf{x}$ in $\Sigma_{\mathbf{a}}$ and called $\mathbf{x}+\mathbf{a}$ in Σ_0. Then equations (4.4) and (1) give

$$\langle\mathbf{x},\mu;0|U(-\mathbf{a})|\psi\rangle = \psi_\mu(\mathbf{x}+\mathbf{a};0) = \psi_\mu(\mathbf{x};\mathbf{a}), \tag{2}$$

where again $|\mathbf{x},\mu;0\rangle$ indicates the state in which the system is located at the point labelled by $\mathbf{x}$ in Σ_0. This equation tells us that $|\psi\rangle$ has the same wavefunction in $\Sigma_{\mathbf{a}}$ that $|\bar\psi\rangle \equiv U(-\mathbf{a})|\psi\rangle$ has in Σ_0. Therefore, moving the origin of our coordinates through a vector $\mathbf{a}$ has the same effect on an arbitrary state's wavefunction as moving the system itself through $-\mathbf{a}$. Physically moving the system is known as an **active transformation**, whereas leaving the state alone but changing the coordinate system is called a **passive transformation**. The infinitesimal vectors required to make logically equivalent active and passive transformations differ in sign. This sign difference reflects the fact that if you move backwards, the world around you seems to move forwards; hence moving the origin of one's coordinates back by $\delta\mathbf{a}$ has the same effect as moving the system forward by $\delta\mathbf{a}$. In this book we confine ourselves to active transformations.

posed above as to what an observable's operator actually does in the case of the momentum operators.

Equation (2.78) enables us to expand a state of well-defined position $\mathbf{x}_0$ in terms of momentum eigenstates. We have

$$|\mathbf{x}_0,\mu\rangle = \int \mathrm{d}^3\mathbf{p}\,|\mathbf{p},\mu\rangle\langle\mathbf{p},\mu|\mathbf{x}_0,\mu\rangle = \frac{1}{h^{3/2}}\int \mathrm{d}^3\mathbf{p}\,\mathrm{e}^{-\mathrm{i}\mathbf{x}_0\cdot\mathbf{p}/\hbar}|\mathbf{p},\mu\rangle. \tag{4.6}$$

Applying the translation operator, we obtain with (4.3)

$$\begin{aligned} U(\mathbf{a})|\mathbf{x}_0,\mu\rangle &= \frac{1}{h^{3/2}}\int \mathrm{d}^3\mathbf{p}\,\mathrm{e}^{-\mathrm{i}\mathbf{x}_0\cdot\mathbf{p}/\hbar}\,U(\mathbf{a})|\mathbf{p},\mu\rangle \\ &= \frac{1}{h^{3/2}}\int \mathrm{d}^3\mathbf{p}\,\mathrm{e}^{-\mathrm{i}(\mathbf{x}_0+\mathbf{a})\cdot\mathbf{p}/\hbar}\,|\mathbf{p},\mu\rangle \\ &= |\mathbf{x}_0+\mathbf{a},\mu\rangle, \end{aligned} \tag{4.7}$$

which is a new state, in which the system is definitely located at $\mathbf{x}_0+\mathbf{a}$, as we would expect.

4.1.2 Continuous transformations and generators

In §1.3 we saw that the normalisation condition $\langle\psi|\psi\rangle = 1$ expresses the fact that when we make a measurement of any observable, we will measure

Box 4.2: Operators from expectation values

In this box we show that if

$$\langle\psi|A|\psi\rangle = \langle\psi|B|\psi\rangle, \tag{1}$$

for every state $|\psi\rangle$, then the operators A and B are identical. We set $|\psi\rangle = |\phi\rangle + \lambda|\chi\rangle$, where λ is a complex number. Then equation (1) implies

$$\lambda\left(\langle\phi|A|\chi\rangle - \langle\phi|B|\chi\rangle\right) = \lambda^*\left(\langle\chi|B|\phi\rangle - \langle\chi|A|\phi\rangle\right). \tag{2}$$

Since equation (1) is valid for any state $|\psi\rangle$, equation (2) remains valid as we vary λ. If the coefficients of λ and λ^* are non-zero, we can cause the left and right sides of (2) to change differently by varying the phase of λ; they can be equal irrespective of the phase of λ only if the coefficients vanish. This shows that $\langle\chi|A|\phi\rangle = \langle\chi|B|\phi\rangle$ for for arbitrary states $|\phi\rangle$ and $|\chi\rangle$, from which it follows that $A = B$.

some value; for example, if we determine the system's location, we will find it *somewhere*. The normalisation condition must be unaffected by any transformation that we make on a system, so the transformation operator[1] U must have the property that for any state $|\psi\rangle$

$$1 = \langle\psi'|\psi'\rangle = \langle\psi|U^\dagger U|\psi\rangle. \tag{4.8}$$

From this requirement we can infer by the argument given in Box 4.2 (with $A = U^\dagger U$ and $B = I$, the identity operator) that $U^\dagger U = I$, so $U^\dagger = U^{-1}$. Operators with this property are called **unitary operators**. When we transform all states with a unitary operator, we leave unchanged all amplitudes: $\langle\phi'|\psi'\rangle = \langle\phi|\psi\rangle$ for any states $|\phi\rangle$ and $|\psi\rangle$.

Exactly how we construct a unitary operator depends on the type of transformation we wish it to make. The identity operator is the unitary operator that represents doing nothing to our system. The translation operator $U(\mathbf{a})$ can be made to approach the identity as closely as we please by diminishing the magnitude of $\mathbf{a}$. Many other unitary operators also have a parameter θ that can be reduced to zero such that the operator tends to the identity. In this case we can write for small $\delta\theta$

$$U(\delta\theta) = I - \mathrm{i}\delta\theta\,\tau + \mathrm{O}(\delta\theta)^2, \tag{4.9}$$

where the factor of i is a matter of convention and τ is an operator. The unitarity of U implies that

$$I = U^\dagger(\delta\theta)U(\delta\theta) = I + \mathrm{i}\delta\theta\,(\tau^\dagger - \tau) + \mathrm{O}(\delta\theta^2). \tag{4.10}$$

Equating powers of $\delta\theta$ on the two sides of the equation, we deduce that τ is Hermitian, so it may be an observable. If so, its eigenkets are states in which the system has well-defined values of the observable τ.

We obtain an important equation by using equation (4.9) to evaluate $|\psi'\rangle \equiv U(\delta\theta)|\psi\rangle$. Subtracting $|\psi\rangle$ from both sides of the resulting equation, dividing through by $\delta\theta$ and proceeding to the limit $\delta\theta \to 0$, we obtain

$$\mathrm{i}\frac{\partial|\psi'\rangle}{\partial\theta} = \tau|\psi'\rangle. \tag{4.11}$$

Thus the observable τ gives the rate at which $|\psi\rangle$ changes when we increase the parameter θ in the unitary transformation that τ generates. Equation (4.5) is a concrete example of this equation in action.

---

[1] We restrict ourselves to the case in which the operator U is linear, as is every operator used in this book. In consequence, we are unable to consider time reversal.

A finite transformation can be generated by repeatedly performing an infinitesimal one. Specifically, if we transform N times with $U(\delta\theta)$ with $\delta\theta = \theta/N$, then in the limit $N \to \infty$ we have

$$U(\theta) \equiv \lim_{N\to\infty} \left(1 - \mathrm{i}\frac{\theta}{N}\tau\right)^N = \mathrm{e}^{-\mathrm{i}\theta\tau}. \tag{4.12}$$

This relation is clearly a generalisation of the definition (4.3) of the translation operator. The Hermitian operator τ is called the **generator** of both the unitary operator U and the transformations that U accomplishes; for example, $\mathbf{p}/\hbar$ is the generator of translations.

4.1.3 The rotation operator

Consider what happens if we rotate the system. Whereas in §4.1.1 we constructed a state $|\psi'\rangle = U(\mathbf{a})|\psi\rangle$ that differed from the state $|\psi\rangle$ only in a shift by $\mathbf{a}$ in the location of the centre of mass, we now wish to find a rotation operator that constructs the state $|\psi'\rangle$ that we would get if we could somehow rotate the apparatus on a turntable without disturbing its internal structure in any way. Whereas the orientation of the system is unaffected by a translation, it will be changed by the rotation operator, as is physically evident if we imagine turning a non-spherical object on a turntable.

From §4.1.2 we know that a rotation operator will be unitary, and have a Hermitian generator. Actually, we expect there to be several generators, just as there are three generators, $p_x/\hbar$, $p_y/\hbar$ and $p_z/\hbar$, of translations. Because there are three generators of translations, three numbers, the components of the vector $\mathbf{a}$ in equation (4.3), are required to specify a particular translation. Hence we anticipate that the number of generators of rotations will equal the number of angles that are required to specify a rotation. Two angles are required to specify the axis of rotation, and a third is required to specify the angle through which we rotate. Thus by analogy with equation (4.3), we expect that a general rotation operator can be obtained by exponentiating a linear combination of three generators of rotations, and we write

$$U(\boldsymbol{\alpha}) = \exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}). \tag{4.13}$$

Here $\boldsymbol{\alpha}$ is a vector that specifies a rotation through an angle $|\boldsymbol{\alpha}|$ around the direction of the unit vector $\hat{\boldsymbol{\alpha}}$, and $\mathbf{J}$ is comprised of three Hermitian operators, J_x, J_y and J_z. In the course of this chapter and the next it will become clear that the observable associated with $\mathbf{J}$ is angular momentum. Consequently, the components of $\mathbf{J}$ are called the **angular-momentum operators**.

The role that the angular momentum operators play in rotating the system around the axis $\hat{\boldsymbol{\alpha}}$ is expressed by rewriting equation (4.11) with appropriate substitutions as

$$\mathrm{i}\frac{\partial|\psi\rangle}{\partial\alpha} = \hat{\boldsymbol{\alpha}}\cdot\mathbf{J}|\psi\rangle. \tag{4.14}$$

4.1.4 Discrete transformations

Not all transformations are continuous. In physics, the most prominent example of a discrete transformation is the **parity transformation** $\mathcal{P}$, which swaps the sign of the coordinates of all spatial points; the action of $\mathcal{P}$ on coordinates is represented by the matrix

$$\mathcal{P} \equiv \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{so} \quad \mathcal{P}\mathbf{x} = -\mathbf{x}. \tag{4.15}$$

Notice that $\det \mathcal{P} = -1$, whereas a rotation matrix has $\det \mathbf{R} = +1$. In fact, any linear transformation with determinant equal to -1 can be written as a product of $\mathcal{P}$ and a rotation $\mathbf{R}$.

Let an arbitrary quantum state $|\psi\rangle$ have wavefunction $\psi_\mu(\mathbf{x}) = \langle \mathbf{x}, \mu|\psi\rangle$, where the label μ is the usual shorthand for the system's orientation. Then the quantum parity operator P is defined by

$$\psi'_\mu(\mathbf{x}) \equiv \langle \mathbf{x}, \mu|P|\psi\rangle \equiv \psi_\mu(\mathcal{P}\mathbf{x}) = \psi_\mu(-\mathbf{x}) = \langle -\mathbf{x}, \mu|\psi\rangle. \tag{4.16}$$

The wavefunction of the new state, $|\psi'\rangle = P|\psi\rangle$, takes the same value at $\mathbf{x}$ that the old wavefunction does at $-\mathbf{x}$. Thus, when the system is in the state $P|\psi\rangle$, it has the same amplitude to be at $\mathbf{x}$ as it had to be at $-\mathbf{x}$ when it was in the state $|\psi\rangle$. The orientation and internal properties of the system are unaffected by P. The invariance of orientation under a parity transformation is not self evident, but in §4.2 we shall see that it follows from the rules that govern commutation of P with $\mathbf{x}$ and $\mathbf{J}$.

Applying the parity operator twice creates a state $|\psi''\rangle = P|\psi'\rangle = P^2|\psi\rangle$ with wavefunction

$$\begin{aligned}\psi''_\mu(\mathbf{x}) &= \langle \mathbf{x}, \mu|P|\psi'\rangle = \langle -\mathbf{x}, \mu|\psi'\rangle = \langle -\mathbf{x}, \mu|P|\psi\rangle = \langle \mathbf{x}, \mu|\psi\rangle \\ &= \psi_\mu(\mathbf{x}).\end{aligned} \tag{4.17}$$

Hence $P^2 = 1$ and an even number of applications of the parity operator leaves the wavefunction unchanged. It also follows that $P = P^{-1}$ is its own inverse.

Because the coordinate reflection $\mathbf{x} \to \mathcal{P}\mathbf{x}$ involves no free parameters that could be taken to be infinitesimal, P is not associated with a Hermitian generator. However, we can show that P is itself Hermitian:

$$\begin{aligned}\langle \phi|P|\psi\rangle^* &= \int \mathrm{d}^3\mathbf{x} \sum_\mu (\langle \phi|\mathbf{x}, \mu\rangle\langle \mathbf{x}, \mu|P|\psi\rangle)^* \\ &= \int \mathrm{d}^3\mathbf{x} \sum_\mu (\langle \phi|\mathbf{x}, \mu\rangle\langle -\mathbf{x}, \mu|\psi\rangle)^* \\ &= \int \mathrm{d}^3\mathbf{x} \sum_\mu \langle \psi| -\mathbf{x}, \mu\rangle\langle \mathbf{x}, \mu|P^2|\phi\rangle \\ &= \int \mathrm{d}^3\mathbf{x} \sum_\mu \langle \psi| -\mathbf{x}, \mu\rangle\langle -\mathbf{x}, \mu|P|\phi\rangle = \langle \psi|P|\phi\rangle,\end{aligned} \tag{4.18}$$

so $P^\dagger = P$. P is also unitary because $P^{-1} = P = P^\dagger$. Hence from the discussion of §4.1.2 it follows that transforming all states with P will preserve all amplitudes for the system.

Suppose now that $|P\rangle$ is an eigenket of P, with eigenvalue λ. Then $|P\rangle = P^2|P\rangle = \lambda P|P\rangle = \lambda^2|P\rangle$, so $\lambda^2 = 1$. Thus the eigenvalues of P are ± 1. Eigenstates of P are said to have **definite parity**, with $|+\rangle = P|+\rangle$ being a state of **even parity** and $|-\rangle = -P|-\rangle$ being one of **odd parity**.

In §3.1 we found that the stationary-state wavefunctions of a harmonic oscillator are even functions of x when the quantum number n is even, and odd functions of x otherwise. It is clear that these stationary states are also eigenstates of P, those for $n = 0, 2, 4, \ldots$ having even parity and those for $n = 1, 3, 5, \ldots$ having odd parity.

4.2 Transformations of operators

When we move an object around, we expect to find it in a new place. Specifically, suppose $\langle\psi|\mathbf{x}|\psi\rangle = \mathbf{x}_0$ for some state $|\psi\rangle$. Since $\mathbf{x}_0$ just labels a spatial point, it must behave under translations and rotations like any vector. For example, translating a system that is in the state $|\psi\rangle$ through $\mathbf{a}$, we obtain a new state $|\psi'\rangle$ which has $\langle\psi'|\mathbf{x}|\psi'\rangle = \mathbf{x}_0 + \mathbf{a} = \langle\psi|\mathbf{x} + I\mathbf{a}|\psi\rangle$. On the other hand, from §4.1.1 we know that $\langle\psi'|\mathbf{x}|\psi'\rangle = \langle\psi|U^\dagger(\mathbf{a})\mathbf{x}U(\mathbf{a})|\psi\rangle$. Since these expectation values must be equal for any initial state $|\psi\rangle$, it follows from the argument given in Box 4.2 that

$$U^\dagger(\mathbf{a})\,\mathbf{x}\,U(\mathbf{a}) = \mathbf{x} + \mathbf{a}, \tag{4.19}$$

where the identity operator is understood to multiply the constant $\mathbf{a}$. For an infinitesimal translation with $\mathbf{a} \to \delta\mathbf{a}$ we have $U(\mathbf{a}) \simeq 1 - \mathrm{i}\mathbf{a}\cdot\mathbf{p}/\hbar$. So

$$\begin{aligned} \mathbf{x} + \delta\mathbf{a} &\simeq \left(1 + \mathrm{i}\frac{\delta\mathbf{a}\cdot\mathbf{p}}{\hbar}\right)\,\mathbf{x}\,\left(1 - \mathrm{i}\frac{\delta\mathbf{a}\cdot\mathbf{p}}{\hbar}\right) \\ &= \mathbf{x} - \frac{\mathrm{i}}{\hbar}\,[\mathbf{x}, \delta\mathbf{a}\cdot\mathbf{p}] + \mathrm{O}(\delta\mathbf{a})^2. \end{aligned} \tag{4.20}$$

For this to be true for all small vectors $\delta\mathbf{a}$, $\mathbf{x}$ and $\mathbf{p}$ must satisfy the commutation relations

$$[x_i, p_j] = \mathrm{i}\hbar\delta_{ij} \tag{4.21}$$

in accordance with equation (2.54). Here we see that this commutation relation arises as a natural consequence of the properties of $\mathbf{x}$ under translations. For a finite translation, we can write

$$U^\dagger(\mathbf{a})\,\mathbf{x}\,U(\mathbf{a}) = U^\dagger(\mathbf{a})U(\mathbf{a})\mathbf{x} + U^\dagger(\mathbf{a})[\mathbf{x}, U(\mathbf{a})] = \mathbf{x} + U^\dagger(\mathbf{a})\,[\mathbf{x}, U(\mathbf{a})]\,. \tag{4.22}$$

We use equation (2.25) to evaluate the commutator on the right. Treating U as the function $\mathrm{e}^{-\mathrm{i}\mathbf{a}\cdot\mathbf{p}/\hbar}$ of $\mathbf{a}\cdot\mathbf{p}$, we find

$$U^\dagger(\mathbf{a})\,\mathbf{x}\,U(\mathbf{a}) = \mathbf{x} - \frac{\mathrm{i}}{\hbar}U^\dagger(\mathbf{a})[\mathbf{x}, \mathbf{a}\cdot\mathbf{p}]U(\mathbf{a}) = \mathbf{x} + \mathbf{a} \tag{4.23}$$

as equation (4.19) requires.

Similarly, under rotations ordinary spatial vectors have components which transform as $\mathbf{v} \to \mathbf{R}(\boldsymbol{\alpha})\mathbf{v}$, where $\mathbf{R}(\boldsymbol{\alpha})$ is a matrix describing a rotation through angle $|\boldsymbol{\alpha}|$ around the $\hat{\boldsymbol{\alpha}}$ axis. The expectation values $\langle\psi|\mathbf{x}|\psi\rangle = \mathbf{x}_0$ should then transform in this way. In §4.1.2 we saw that when a system is rotated through an angle $|\boldsymbol{\alpha}|$ around the $\hat{\boldsymbol{\alpha}}$ axis, its ket $|\psi\rangle$ should be multiplied by $U(\boldsymbol{\alpha}) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}}$. If this transformation of $|\psi\rangle$ is to be consistent with the rotation of the expectation value of $\mathbf{x}$, we need

$$\mathbf{R}(\boldsymbol{\alpha})\langle\psi|\mathbf{x}|\psi\rangle = \langle\psi'|\mathbf{x}|\psi'\rangle = \langle\psi|U^\dagger(\boldsymbol{\alpha})\,\mathbf{x}\,U(\boldsymbol{\alpha})|\psi\rangle. \tag{4.24}$$

Since this must hold for any state $|\psi\rangle$, from the argument given in Box 4.2 it follows that

$$\mathbf{R}(\boldsymbol{\alpha})\mathbf{x} = U^\dagger(\boldsymbol{\alpha})\,\mathbf{x}\,U(\boldsymbol{\alpha}). \tag{4.25}$$

For an infinitesimal rotation, $\boldsymbol{\alpha} \to \delta\boldsymbol{\alpha}$ and $\mathbf{R}(\boldsymbol{\alpha})\mathbf{x} \simeq \mathbf{x} + \delta\boldsymbol{\alpha}\times\mathbf{x}$ as is shown in Box 4.3, so equation (4.25) becomes

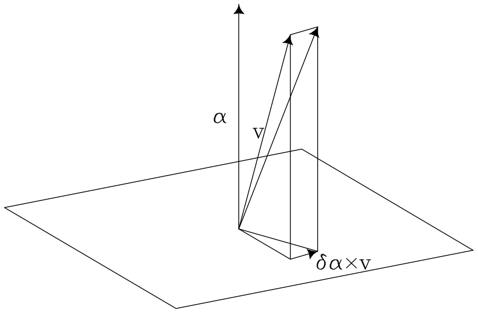
$$\begin{aligned} \mathbf{x} + \delta\boldsymbol{\alpha}\times\mathbf{x} &\simeq (1 + \mathrm{i}\delta\boldsymbol{\alpha}\cdot\mathbf{J})\;\mathbf{x}\;(1 - \mathrm{i}\delta\boldsymbol{\alpha}\cdot\mathbf{J}) \\ &= \mathbf{x} + \mathrm{i}\,[\delta\boldsymbol{\alpha}\cdot\mathbf{J}, \mathbf{x}] + \mathrm{O}(\delta\boldsymbol{\alpha})^2. \end{aligned} \tag{4.26}$$

In components, the vector product $\delta\boldsymbol{\alpha}\times\mathbf{x}$ can be written

$$(\delta\boldsymbol{\alpha}\times\mathbf{x})_i = \sum_{jk}\epsilon_{ijk}\delta\alpha_j\,x_k, \tag{4.27}$$

Box 4.3: Rotations in ordinary space

A rotation matrix $\mathbf{R}$ is defined by the conditions $\mathbf{R}^{\mathrm{T}} = \mathbf{R}^{-1}$ and $\det(\mathbf{R}) = +1$. If $\mathbf{R}(\boldsymbol{\alpha})$ rotates around the $\hat{\boldsymbol{\alpha}}$ axis, it should leave this axis invariant so $\mathbf{R}(\boldsymbol{\alpha})\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}$. For a rotation through an angle $|\boldsymbol{\alpha}|$, $\mathrm{Tr}\,\mathbf{R}(\boldsymbol{\alpha}) = 1 + 2\cos|\boldsymbol{\alpha}|$.



Let $\boldsymbol{\alpha}$ be an infinitesimal rotation vector: that is, a rotation through α around the axis that is in the direction of the unit vector $\hat{\boldsymbol{\alpha}}$. We consider the effect of rotating an arbitrary vector $\mathbf{v}$ through angle $\boldsymbol{\alpha}$. The component of $\mathbf{v}$ parallel to $\boldsymbol{\alpha}$ is unchanged by the rotation. The figure shows the projection of $\mathbf{v}$ into the plane perpendicular to $\boldsymbol{\alpha}$. The rotated vector is seen to be the vectorial sum of $\mathbf{v}$ and the infinitesimal vector $\boldsymbol{\alpha} \times \mathbf{v}$ that joins the end of $\mathbf{v}$ before and after rotation. That is

$$\mathbf{v}' = \mathbf{v} + \boldsymbol{\alpha} \times \mathbf{v}.$$

where ϵ_{ijk} is the object that changes sign if any two subscripts are interchanged and has $\epsilon_{xyz} = 1$ (Appendix A). For example, equation (4.27) gives $(\delta\boldsymbol{\alpha} \times \mathbf{x})_x = \sum_{jk} \epsilon_{xjk}\delta\alpha_j\, x_k = \epsilon_{xyz}\delta\alpha_y\, x_z + \epsilon_{xzy}\delta\alpha_z\, x_y = \delta\alpha_y z - \delta\alpha_z y$. The i^{th} components of equation (4.26) is

$$\sum_{jk} \epsilon_{ijk}\delta\alpha_j\, x_k = \mathrm{i}\sum_j \delta\alpha_j [J_j, x_i]. \tag{4.28}$$

Since this equation holds for arbitrary $\delta\boldsymbol{\alpha}$, we conclude that the position and angular momentum operators x_i and J_j must satisfy the commutation relation

$$[J_i, x_j] = \mathrm{i}\sum_k \epsilon_{ijk} x_k. \tag{4.29}$$

In particular, $[J_x, y] = \mathrm{i}z$ and $[J_z, x] = -\mathrm{i}y$, while $[J_x, x] = 0$.

In fact, if the expectation value of *any* operator $\mathbf{v}$ is a spatial vector, then the argument just given in the case of $\mathbf{x}$ shows that the components v_i must satisfy

$$[J_i, v_j] = \mathrm{i}\sum_k \epsilon_{ijk} v_k. \tag{4.30}$$

For example, since momentum is a vector, equation (4.30) with $\mathbf{v} = \mathbf{p}$ gives the commutation relations of $\mathbf{p}$ with $\mathbf{J}$,

$$[J_i, p_j] = \mathrm{i}\sum_k \epsilon_{ijk} p_k. \tag{4.31}$$

The product $\boldsymbol{\alpha}\cdot\mathbf{J}$ must be invariant under coordinate rotations because the operator $U(\boldsymbol{\alpha}) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}}$ depends on the *direction* $\hat{\boldsymbol{\alpha}}$ and not on the numbers used to quantify that direction. Since $\boldsymbol{\alpha}$ is an arbitrary vector, the invariance of $\boldsymbol{\alpha}\cdot\mathbf{J}$ under rotations implies that under rotations the components of $\mathbf{J}$ transform like those of a vector. Hence, in equation (4.30) we can replace $\mathbf{v}$ by $\mathbf{J}$ to obtain the commutation relation

$$[J_i, J_j] = \mathrm{i}\sum_k \epsilon_{ijk} J_k. \tag{4.32}$$

In §7.1 shall deduce the spectrum of the angular-momentum operators from this relation.

We now show that $\mathbf{J}$ commutes with any operator S whose expectation value is a scalar. The proof is simple: $\langle\psi|S|\psi\rangle$, being a scalar, is not affected by rotations, so

$$\langle\psi'|S|\psi'\rangle = \langle\psi|U^\dagger(\boldsymbol{\alpha})\,S\,U(\boldsymbol{\alpha})|\psi\rangle = \langle\psi|S|\psi\rangle. \tag{4.33}$$

Equating the operators on either side of the second equality and using $U^{-1} = U^\dagger$ we have $[S,U]=0$. Restricting U to an infinitesimal rotation gives

$$S \simeq (1+\mathrm{i}\delta\boldsymbol{\alpha}\cdot\mathbf{J})\;S\;(1-\mathrm{i}\delta\boldsymbol{\alpha}\cdot\mathbf{J}) = S + \mathrm{i}\delta\boldsymbol{\alpha}\cdot[\mathbf{J},S] + \mathrm{O}(\delta\boldsymbol{\alpha})^2. \tag{4.34}$$

Since $\delta\boldsymbol{\alpha}$ is arbitrary, it follows that

$$[\mathbf{J},S]=0. \tag{4.35}$$

Among other things, this tells us that $[\mathbf{J},\mathbf{x}\cdot\mathbf{x}] = [\mathbf{J},\mathbf{p}\cdot\mathbf{p}] = [\mathbf{J},\mathbf{x}\cdot\mathbf{p}] = 0$. It is straightforward to check that these results are consistent with the vector commutation relations (4.30) (Problem 4.1). It also follows that $J^2 = \mathbf{J}\cdot\mathbf{J}$ commutes with all of the J_i,

$$[\mathbf{J},J^2]=0. \tag{4.36}$$

Equations (4.32) and (4.36) imply that it is possible to find a complete set of simultaneous eigenstates for both J^2 and any one component of $\mathbf{J}$ (but only one).

Under a parity transform, coordinates behave as $\mathbf{x}\to\mathcal{P}\mathbf{x}=-\mathbf{x}$ whereas quantum states transform as $|\psi\rangle\to|\psi'\rangle = P|\psi\rangle$, so

$$-\langle\psi|\mathbf{x}|\psi\rangle = \mathcal{P}\langle\psi|\mathbf{x}|\psi\rangle = \langle\psi'|\mathbf{x}|\psi'\rangle = \langle\psi|P^\dagger\mathbf{x}P|\psi\rangle, \tag{4.37}$$

which implies that $P^\dagger\mathbf{x}P=-\mathbf{x}$ or, since P is a unitary operator,

$$\{\mathbf{x},P\}\equiv\mathbf{x}P+P\mathbf{x}=0. \tag{4.38}$$

Two operators A and B for which $\{A,B\}=0$ are said to **anticommute**, with $\{A,B\}$ being their **anticommutator**. The argument we have just given for $\mathbf{x}$ works with $\mathbf{x}$ replaced by any vector operator $\mathbf{v}$, so we always have

$$\{\mathbf{v},P\}=\mathbf{v}P+P\mathbf{v}=0. \tag{4.39}$$

This relation contains important information about the action of P. Suppose $|\boldsymbol{\omega}\rangle$ is an eigenstate of a vector operator $\mathbf{v}$ with eigenvalues $\boldsymbol{\omega}$ such that $\mathbf{v}|\boldsymbol{\omega}\rangle = \boldsymbol{\omega}|\boldsymbol{\omega}\rangle$. From (4.39) we see that

$$\mathbf{v}|\boldsymbol{\omega}'\rangle = \mathbf{v}\,(P|\boldsymbol{\omega}\rangle) = -P\mathbf{v}|\boldsymbol{\omega}\rangle = -\boldsymbol{\omega}P|\boldsymbol{\omega}\rangle = -\boldsymbol{\omega}|\boldsymbol{\omega}'\rangle \tag{4.40}$$

so the parity-reversed state $|\boldsymbol{\omega}'\rangle = P|\boldsymbol{\omega}\rangle$ is also an eigenstate of $\mathbf{v}$, but the eigenvalue has changed sign.

Let $|\pm\rangle$ be states of definite parity such that $P|\pm\rangle = \pm|\pm\rangle$. With equation (4.39) we deduce that

$$-\langle\pm|\mathbf{v}|\pm\rangle = \mathcal{P}\langle\pm|\mathbf{v}|\pm\rangle = \langle\pm|P^\dagger\mathbf{v}P|\pm\rangle = (\pm)^2\langle\pm|\mathbf{v}|\pm\rangle. \tag{4.41}$$

Since zero is the only number that is equal to minus itself, all vector operators have vanishing expectation value in states of definite parity. More generally, if $|\phi\rangle$ and $|\chi\rangle$ both have the same definite parity, equation (4.39) implies that $\langle\phi|\mathbf{v}|\chi\rangle=0$. We'll use this result in Chapter 9.

In classical physics, a vector product $\mathbf{a}\times\mathbf{b}$ is a **pseudovector**; it behaves like an ordinary vector under rotations, but is invariant under parity, since both $\mathbf{a}$ and $\mathbf{b}$ change sign. We now show that expectation values of the angular momentum operators, $\langle\mathbf{J}\rangle$, are pseudovectors. If v_i are components of a vector operator, then combining equations (4.30) and (4.39), we obtain

$$\{P,[v_i,J_j]\} = \mathrm{i}\sum_k \epsilon_{ijk}\{P,v_k\} = 0. \tag{4.42}$$

We use the identity (4.69) proved in Problem 4.8 to rewrite the left side of this equation. We obtain

$$0 = \{P,[v_i,J_j]\} = [\{P,v_i\},J_j] - \{[P,J_j],v_i\} = -\{[P,J_j],v_i\}. \tag{4.43}$$

Hence the operator $[P,J_j]$ anticommutes with any component of an arbitrary vector. Since P is defined to have precisely this property, $[P,J_j]$ must be proportional to P, that is

$$[P,J_j] = \lambda P, \tag{4.44}$$

where λ must be the same for all values of j because the three coordinate directions are equivalent. Under rotations, the left side transforms like a vector, while the right side is invariant. This is possible only if both sides vanish. Hence the parity operator commutes with all three angular-momentum operators. It now follows that

$$\langle\psi'|\mathbf{J}|\psi'\rangle = \langle\psi|P^\dagger\mathbf{J}P|\psi\rangle = \langle\psi|\mathbf{J}|\psi\rangle, \tag{4.45}$$

so $\langle\mathbf{J}\rangle$ is unchanged by a parity transformation, and is a pseudovector.

4.3 Symmetries and conservation laws

Time changes states: in a given time interval t, the natural evolution of the system causes any state $|\psi,0\rangle$ to evolve to another state $|\psi,t\rangle$. Equation (2.32) gives an explicit expression for $|\psi,t\rangle$. It is easy to see that with the present notation this rule can be written

$$|\psi,t\rangle = \mathrm{e}^{-\mathrm{i}Ht/\hbar}|\psi,0\rangle, \tag{4.46}$$

where H is the Hamiltonian. The time-evolution operator

$$U(t) \equiv \mathrm{e}^{-\mathrm{i}Ht/\hbar} \tag{4.47}$$

is unitary, as we would expect.[2]

Now suppose that the generator τ of some displacement (a translation, a rotation, or something similar) commutes with H. Since these operators commute, their exponentials $U(\theta)$ (eq. 4.12) and $U(t)$ also commute. Consequently, for any state $|\psi\rangle$

$$U(\theta)U(t)|\psi\rangle = U(t)U(\theta)|\psi\rangle. \tag{4.48}$$

The left side is the state you get by waiting for time t and then displacing, while the right side is the state obtained by displacing first and then waiting.

---

[2] The similarity between equations (4.47) and the formula (4.12) for a general unitary transformation suggests that H is the generator of transformations in time. This is not quite true. If we were to push the system forward in time in the same way that we translate it in x, we would delay the instant at which we would impose some given initial conditions, with the result that it would be *less* evolved at a given time t. The time-evolution operator, by contrast, makes the system older. Hence H is the generator of transformations backwards in time.

So the equation says that the system evolves in the same way no matter where you put it. That is, there is a connection between commuting observables and invariance of the physics under displacements. Moreover, in §2.2.1 we saw that when any operator Q commutes with the Hamiltonian, the expectation value of any function of Q is a conserved quantity, and that in consequence, a system that is initially in an eigenstate $|q_i\rangle$ of Q remains in that eigenstate. So whenever the physics is unchanged by a displacement, there is a conserved quantity.

If $[p_x, H] = 0$, this argument implies that the system evolves in the same way wherever it is located. We say that the Hamiltonian is **translationally** invariant. It is a fundamental premise of physics that empty space is the same everywhere, so the Hamiltonian of every isolated system is translationally invariant. Consequently, when a system is isolated, the expectation value of any function of the momentum operators is a conserved quantity, and, if the system is started in a state of well-defined momentum, it will stay in that state. This is Newton's first law.

If $[J_z, H] = 0$, we say that the Hamiltonian is rotationally invariant around the z axis, and our argument implies that the system evolves in the same way no matter how it is turned around the z axis. The expectation value of any function of J_z is constant, and if the state is initially in an eigenstate of J_z with eigenvalue m, it will remain in that state. Consequently, m is a good quantum number. In classical physics invariance of a system's dynamics under rotations around the z axis is associated with conservation of the z component of the system's angular momentum. This fact inspires the identification of $\hbar\mathbf{J}$ with angular momentum.

In general, finding all the operators that commute with a given Hamiltonian is a very difficult problem. However, it is sometimes possible to deduce conserved quantities by direct inspection. For example, the Hamiltonian for a system of n particles that interact with each other, but not with anything else, is

$$H = \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} + \sum_{i<j} V(\mathbf{x}_i - \mathbf{x}_j), \tag{4.49}$$

where the potential-energy function V only depends on the relative positions of the individual particles. Such a Hamiltonian is invariant under translations of all particles together (shifts of the centre of mass coordinate) and thus the total momentum $\mathbf{p}_{\rm tot} = \sum_i \mathbf{p}_i$ of this system is conserved.

If the Hamiltonian is a scalar, then $[H, \mathbf{J}] = 0$, $[H, J^2] = 0$ and $[H, P] = 0$ (Problem 4.9), which implies conservation of angular momentum around any axis, conservation of total angular momentum, and conservation of parity. We have already seen that $[\mathbf{J}, J^2] = [\mathbf{J}, P] = 0$, so for a scalar Hamiltonian we can find complete sets of simultaneous eigenkets of H, P, J^2, and any one of the components of $\mathbf{J}$.

The equation $[H, P] = 0$ implies that if you set up a system that at $t = 0$ is a mirror image of a given system, it will evolve in exactly the same way as the given system. When the evolution of the mirrored system is watched, it will appear identical to the evolution of the given system when the latter is observed in a mirror. Hence, when $[H, P] = 0$, it is impossible to tell whether a system that is being observed, is being watched directly or through a mirror. One of the major surprises of 20th century physics was an experiment by Wu et al.[3] in 1957, which showed that you can see things in a mirror that cannot happen in the real world! That is, there are Hamiltonians for which $[H, P] \neq 0$.

---

[3] Wu, C.S., Ambler, E., Hayward, R.W., Hoppes, D.D. & Hudson, R.P., 1957, Phys. Rev., 105, 1413

4.4 The Heisenberg picture

All physical predictions are extracted from the formalism of quantum mechanics by operating with a bra on a ket to extract a complex number: we either calculate the amplitude for some event as $A = \langle\phi|\psi\rangle$ or the expectation value of a observable through $\langle Q\rangle = \langle\psi|Q\psi\rangle$, where $|Q\psi\rangle \equiv Q|\psi\rangle$. In general our predictions are time-dependent because the state of our system evolves in time according to

$$|\psi, t\rangle = U(t)|\psi, 0\rangle, \tag{4.50}$$

where the time-evolution operator U is defined by equation (4.47).

With every operator of interest we can associate a new time-dependent operator

$$\widetilde{Q}_t \equiv U^\dagger(t)QU(t). \tag{4.51}$$

Then at any time t the expectation value of Q can be written

$$\langle Q\rangle_t = \langle\psi, t|Q|\psi, t\rangle = \langle\psi, 0|U^\dagger(t)QU(t)|\psi, 0\rangle = \langle\psi, 0|\widetilde{Q}_t|\psi, 0\rangle. \tag{4.52}$$

That is, the expectation value at time $t = 0$ of the new operator $\widetilde{Q}_t$ is equal to the expectation value of the original, physical, operator Q at time t. Similarly, when we wish to calculate an amplitude $\langle\phi, t|\psi, t\rangle$ for something to happen at time t, we can argue that on account of the unitarity of $U(t)$ it is equal to a corresponding amplitude at time zero:

$$\langle\phi, t|\psi, t\rangle = \langle\phi, 0|\psi, 0\rangle \quad \text{where} \quad |\phi, t\rangle \equiv U(t)|\phi, 0\rangle. \tag{4.53}$$

Thus if we work with the new time-dependent operators such as $\widetilde{Q}_t$, the only states we require are those at $t = 0$. This formalism, is called the **Heisenberg picture** to distinguish it from the **Schrödinger picture** in which states evolve and operators are normally time-independent.

As we have seen, classical mechanics applies in the limit that it is sufficient to calculate the expectation values of observables, and is concerned with solving the equations of motion of these expectation values. In the Heisenberg picture quantum mechanics is concerned with solving the equations of motion of the time-dependent operators $\widetilde{Q}_t$, etc. Consequently, there is a degree of similarity between the Heisenberg picture and classical mechanics.

It is straightforward to determine the equation of motion of $\widetilde{Q}_t$: we simply differentiate equation (4.51)

$$\frac{\mathrm{d}\widetilde{Q}_t}{\mathrm{d}t} = \frac{\mathrm{d}U^\dagger}{\mathrm{d}t}QU + U^\dagger Q\frac{\mathrm{d}U}{\mathrm{d}t}. \tag{4.54}$$

But differentiating equation (4.47) we have

$$\frac{\mathrm{d}U}{\mathrm{d}t} = -\frac{\mathrm{i}H}{\hbar}U \quad \Rightarrow \quad \frac{\mathrm{d}U^\dagger}{\mathrm{d}t} = \frac{\mathrm{i}H}{\hbar}U^\dagger, \tag{4.55}$$

where we have taken advantage of the fact that U is a function of H and therefore commutes with it. Inserting these expressions into equation (4.54) we obtain

$$\begin{aligned} \mathrm{i}\hbar\frac{\mathrm{d}\widetilde{Q}_t}{\mathrm{d}t} &= -HU^\dagger QU + U^\dagger QUH \\ &= [\widetilde{Q}_t, H]. \end{aligned} \tag{4.56}$$

This result is similar to Ehrenfest's theorem (eq. 2.34) as it has to be because Ehrenfest's theorem must be recovered if we pre- and post-multiply each side by the time-independent state $|\psi, 0\rangle$.

The Heisenberg picture is most widely used in the quantum theory of fields. In this theory one needs essentially only one state, the vacuum in the remote past $|0\rangle$, which we assume was empty. Excitations of the vacuum are interpreted as particles, each mode of excitation being associated with a different type of particle (photons, electron, up-quarks, etc). The theory is concerned with the dynamics of operators that excite the vacuum, creating particles, which then propagate to other locations, where they are detected (annihilated) by similar operators. Sometimes one mode of excitation of the vacuum morphs into one or more different modes of the vacuum, and such an event is interpreted as the decay of one type of particle into other particles. The amplitude for any such sequence of events is obtained as a number of the form $\langle 0|\widetilde{A}_1\widetilde{A}_2\ldots\widetilde{A}_n|0\rangle$, where the operators $\widetilde{A}_i$ are creation or annihilation operators for the appropriate particles in the Heisenberg picture.

4.5 What is the essence of quantum mechanics?

It is sometimes said that commutation relations such as $[x_i, p_j] = \mathrm{i}\hbar\delta_{ij}$ and $[J_i, J_j] = \mathrm{i}\sum_k \epsilon_{ijk}J_k$ are inherently *quantum* mechanical, but this is not true.

Take for example an ordinary classical rotation matrix $\mathbf{R}(\boldsymbol{\alpha})$ which rotates spatial vectors as $\mathbf{v} \to \mathbf{v}' = \mathbf{R}(\boldsymbol{\alpha})\mathbf{v}$. Define matrices $\mathcal{J}_x$, $\mathcal{J}_y$ and $\mathcal{J}_z$ via

$$\exp\left(-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{J}}\right) \equiv \mathbf{R}(\boldsymbol{\alpha}), \tag{4.57}$$

where the exponential of a matrix is defined in terms of the power series for e^x. Clearly, the $\mathcal{J}_i$ must be 3×3 matrices, and, since $\mathbf{R}(\boldsymbol{\alpha})$ is real and the angles $\boldsymbol{\alpha}$ are arbitrary, the $\mathcal{J}_i$ must be pure imaginary. Finally, orthogonality of $\mathbf{R}$ requires

$$\begin{aligned}\mathbf{I} = \mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\mathbf{R}(\boldsymbol{\alpha}) &= \exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{J}})^{\mathrm{T}}\exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{J}})\\ &= \exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{J}}^{\mathrm{T}})\exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{J}}) = \exp\{-\mathrm{i}\boldsymbol{\alpha}\cdot(\boldsymbol{\mathcal{J}}^{\mathrm{T}}+\boldsymbol{\mathcal{J}})\},\end{aligned} \tag{4.58}$$

where the last equality exploits the fact that any matrix commutes with its transpose. Since $\boldsymbol{\alpha}$ is an arbitrary vector, it now follows that $\mathcal{J}_i^{\mathrm{T}} = -\mathcal{J}_i$, so $\mathcal{J}_i$ is antisymmetric. A pure imaginary antisymmetric matrix is a Hermitian matrix. Thus the $\mathcal{J}_i$ are Hermitian.

For any two vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, it is easy to show that the product $\mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\mathbf{R}(\boldsymbol{\beta})\mathbf{R}(\boldsymbol{\alpha})$ is an orthogonal matrix with determinant $+1$, so it is a rotation matrix. It leaves the vector $\boldsymbol{\beta}' \equiv \mathbf{R}(-\boldsymbol{\alpha})\boldsymbol{\beta} = \mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\boldsymbol{\beta}$ invariant:

$$\Big\{\mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\mathbf{R}(\boldsymbol{\beta})\mathbf{R}(\boldsymbol{\alpha})\Big\}\boldsymbol{\beta}' = \mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\mathbf{R}(\boldsymbol{\beta})\boldsymbol{\beta} = \mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\boldsymbol{\beta} = \boldsymbol{\beta}'. \tag{4.59}$$

Hence $\boldsymbol{\beta}'$ is the axis of this rotation. Therefore

$$\mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\mathbf{R}(\boldsymbol{\beta})\mathbf{R}(\boldsymbol{\alpha}) = \mathbf{R}(\boldsymbol{\beta}') = \mathbf{R}(\mathbf{R}(-\boldsymbol{\alpha})\boldsymbol{\beta}). \tag{4.60}$$

In Box 4.3 we showed that when $|\boldsymbol{\alpha}|$ is infinitesimal, $\mathbf{R}(-\boldsymbol{\alpha})\boldsymbol{\beta} \simeq \boldsymbol{\beta} - \boldsymbol{\alpha}\times\boldsymbol{\beta}$, so when $\boldsymbol{\beta}$ is also infinitesimal, equation (4.60) can be written in terms of the classical generators (4.57) as

$$\left(1+\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{J}}\right)\left(1-\mathrm{i}\boldsymbol{\beta}\cdot\boldsymbol{\mathcal{J}}\right)\left(1-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\mathcal{J}}\right) \simeq 1 - \mathrm{i}(\boldsymbol{\beta}-\boldsymbol{\alpha}\times\boldsymbol{\beta})\cdot\boldsymbol{\mathcal{J}}. \tag{4.61}$$

The zeroth order terms (‘1’) and those involving only $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ cancel, but the terms involving both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ cancel only if

$$\alpha_i\beta_j[\mathcal{J}_i, \mathcal{J}_j] = \mathrm{i}\alpha_i\beta_j\sum_k \epsilon_{ijk}\mathcal{J}_k. \tag{4.62}$$

This equation can hold for all directions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ only if the $\mathcal{J}_i$ satisfy

$$[\mathcal{J}_i, \mathcal{J}_j] = \mathrm{i}\sum_k \epsilon_{ijk}\mathcal{J}_k, \tag{4.63}$$

which is identical to the 'quantum' commutation relation (4.32). Our rederivation of these commutation relations from entirely classical considerations is possible because the relations reflect the fact that the order in which you rotate an object around two different axes matters (Problem 4.6). This is a statement about the geometry of space that has to be recognised by both quantum and classical mechanics.

In Appendix C it is shown that in classical statistical mechanics, each component of position, x_i, and momentum, p_i, is associated with a Hermitian operator $\hat{x}_i$ or $\hat{p}_i$ that acts on functions on phase space. The operator $\hat{p}_i$ generates translations along x_i, while $\hat{x}_i$ generates translations along p_i (boosts). The operators $\hat{L}_i$ associated with angular momentum satisfy the commutation relation $[\hat{L}_x, \hat{L}_y] = \mathrm{i}\mathcal{H}\hat{L}_z$, where $\mathcal{H}$ is a number with the same dimensions as $\hbar$ and a magnitude that depends on how $\hat{x}_i$ and $\hat{p}_i$ are normalised.

If the form of the commutation relations is not special to quantum mechanics, what is? In quantum mechanics, complete information about any system is contained in its ket $|\psi\rangle$. There is nothing else. From $|\psi\rangle$ we can evaluate amplitudes such as $\langle\mathbf{x}, \mu|\psi\rangle$ for the system to be found at $\mathbf{x}$ with orientation μ. If we do not care about μ, the total probability for $|\psi\rangle$ to be found at $\mathbf{x}$ is

$$\mathrm{Prob}(\text{at } \mathbf{x}|\psi) = \sum_\mu \big|\langle\mathbf{x}, \mu|\psi\rangle\big|^2. \tag{4.64}$$

Eigenstates of the $\mathbf{x}$ operator with eigenvalue $\mathbf{x}_0$ are states in which the system is definitely at $\mathbf{x}_0$, while eigenstates of the $\mathbf{p}$ operator with eigenvalue $\hbar\mathbf{k}$ are states in which the system definitely has momentum $\hbar\mathbf{k}$.

By contrast, in classical statistical mechanics we declare at the outset that a well defined state is one that has definite values for all measurable quantities, so it has a definite position, momentum, orientation etc. The eigenfunctions of $\hat{\mathbf{p}}$ or $\hat{\mathbf{L}}$ do not represent states of definite momentum or angular momentum, because we have already defined what such states are.

Classical statistical mechanics knows nothing about probability amplitudes, but interprets the functions on phase space on which $\hat{\mathbf{p}}$ or $\hat{\mathbf{L}}$ act as probability distributions. This is possible because, as we show in Appendix C, the integral of such a distribution can be normalised to one and is conserved. We can certainly expand any such distribution in the eigenfunctions of, say $\hat{\mathbf{p}}$. However, as in quantum mechanics the expansion coefficients will not be positive – in fact, they will generally be complex. Hence they cannot be interpreted as probabilities. What makes quantum mechanics fundamentally different is its reliance on complex quantum amplitudes, and the physical interpretation that it gives to a functional expansion through the fundamental rule (1.11) for adding quantum amplitudes. Quantum mechanics is therefore naturally formulated in terms of states $|\psi\rangle$ that inhabit a complex vector space of arbitrary dimension – a so called **Hilbert space**. These states may always be expanded in terms of a complete set of eigenstates of a Hermitian operator, and the (complex) expansion coefficients have a simple physical interpretation.

Classical statistical mechanics is restricted to probabilities, which have to be real, non-negative numbers and are therefore never expansion coefficients. Quantum and classical mechanics incorporate the same commutation relations, however, because, as we stressed in §4.2, these follow from the geometry of space.

Problems

4.1 Verify that $[\mathbf{J},\mathbf{x}\cdot\mathbf{x}]=0$ and $[\mathbf{J},\mathbf{x}\cdot\mathbf{p}]=0$ by using the commutation relations $[x_i,J_j]=\mathrm{i}\sum_k\epsilon_{ijk}x_k$ and $[p_i,J_j]=\mathrm{i}\sum_k\epsilon_{ijk}p_k$.

4.2* Show that the vector product $\mathbf{a}\times\mathbf{b}$ of two classical vectors transforms like a vector under rotations. Hint: A rotation matrix $\mathbf{R}$ satisfies the relations $\mathbf{R}\cdot\mathbf{R}^T=\mathbf{I}$ and $\det(\mathbf{R})=1$, which in tensor notation read $\sum_p R_{ip}R_{tp}=\delta_{it}$ and $\sum_{ijk}\epsilon_{ijk}R_{ir}R_{js}R_{kt}=\epsilon_{rst}$.

4.3* We have shown that $[v_i,J_j]=\mathrm{i}\sum_k\epsilon_{ijk}v_k$ for any operator whose components v_i form a vector. The expectation value of this operator relation in any state $|\psi\rangle$ is then $\langle\psi|[v_i,J_j]|\psi\rangle=\mathrm{i}\sum_k\epsilon_{ijk}\langle\psi|v_k|\psi\rangle$. Check that with $U(\boldsymbol{\alpha})=\mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}}$ this relation is consistent under a further rotation $|\psi\rangle\to|\psi'\rangle=U(\boldsymbol{\alpha})|\psi\rangle$ by evaluating both sides separately.

4.4* The matrix for rotating an ordinary vector by ϕ around the z axis is

$$\mathbf{R}(\phi)\equiv\begin{pmatrix}\cos\phi & -\sin\phi & 0\\ \sin\phi & \cos\phi & 0\\ 0 & 0 & 1\end{pmatrix} \tag{4.65}$$

From $\mathbf{R}$ calculate the matrix $\boldsymbol{\mathcal{J}}_z$ that appears in $\mathbf{R}(\phi)=\exp(-\mathrm{i}\boldsymbol{\mathcal{J}}_z\phi)$. Introduce new coordinates $u_1\equiv(x-\mathrm{i}y)/\sqrt{2}$, $u_2=z$ and $u_3\equiv(x+\mathrm{i}y)/\sqrt{2}$. Write down the matrix $\mathbf{M}$ that appears in $\mathbf{u}=\mathbf{M}\cdot\mathbf{x}$ [where $\mathbf{x}\equiv(x,y,z)$] and show that it is unitary. Then show that

$$\boldsymbol{\mathcal{J}}'_z\equiv\mathbf{M}\cdot\boldsymbol{\mathcal{J}}_z\cdot\mathbf{M}^\dagger. \tag{4.66}$$

is identical with S_z in the set of spin-one Pauli analogues

$$S_x=\frac{1}{\sqrt{2}}\begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix},\quad S_y=\frac{1}{\sqrt{2}}\begin{pmatrix}0&-\mathrm{i}&0\\ \mathrm{i}&0&-\mathrm{i}\\0&\mathrm{i}&0\end{pmatrix},\quad S_z=\begin{pmatrix}1&0&0\\0&0&0\\0&0&-1\end{pmatrix}. \tag{4.67}$$

Write down the matrix $\boldsymbol{\mathcal{J}}_x$ whose exponential generates rotations around the x axis, calculate $\boldsymbol{\mathcal{J}}'_x$ by analogy with equation (4.66) and check that your result agrees with S_x in the set (4.67). Explain as fully as you can the meaning of these calculations.

4.5 Determine the commutator $[\boldsymbol{\mathcal{J}}'_x,\boldsymbol{\mathcal{J}}'_z]$ of the generators used in Problem 4.4. Show that it is equal to $-\mathrm{i}\boldsymbol{\mathcal{J}}'_y$, where $\boldsymbol{\mathcal{J}}'_y$ is identical with S_y in the set (4.67).

4.6* Show that if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are non-parallel vectors, $\boldsymbol{\alpha}$ is not invariant under the combined rotation $\mathbf{R}(\boldsymbol{\alpha})\mathbf{R}(\boldsymbol{\beta})$. Hence show that $\mathbf{R}^{\mathrm{T}}(\boldsymbol{\beta})\mathbf{R}^{\mathrm{T}}(\boldsymbol{\alpha})\mathbf{R}(\boldsymbol{\beta})\mathbf{R}(\boldsymbol{\alpha})$ is not the identity operation. Explain the physical significance of this result.

4.7* In this problem you derive the wavefunction

$$\langle\mathbf{x}|\mathbf{p}\rangle=\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}/\hbar} \tag{4.68}$$

of a state of well defined momentum from the properties of the translation operator $U(\mathbf{a})$. The state $|\mathbf{k}\rangle$ is one of well-defined momentum $\hbar\mathbf{k}$. How would you characterise the state $|\mathbf{k}'\rangle\equiv U(\mathbf{a})|\mathbf{k}\rangle$? Show that the wavefunctions of these states are related by $u_{\mathbf{k}'}(\mathbf{x})=\mathrm{e}^{-\mathrm{i}\mathbf{a}\cdot\mathbf{k}}u_{\mathbf{k}}(\mathbf{x})$ and $u_{\mathbf{k}'}(\mathbf{x})=u_{\mathbf{k}}(\mathbf{x}-\mathbf{a})$. Hence obtain equation (4.68).

4.8 By expanding the anticommutator on the left and then applying the third rule of the set (2.22), show that any three operators satisfy the identity

$$[\{A,B\},C]=\{A,[B,C]\}+\{[A,C],B\}. \tag{4.69}$$

4.9 Let P be the parity operator and S an arbitrary scalar operator. Explain why P and S must commute.

5

Motion in step potentials

We follow up our study of the harmonic oscillator by looking at motion in a wider range of one-dimensional potentials $V(x)$. The potentials we study will be artificial in that they will only vary in sharp steps, but they will enable us to explore analytically some features of quantum mechanics that are generic and hidden from us in the classical limit.

5.1 Square potential well

We look for energy eigenstates of a particle that moves in the potential (Figure 5.1)

$$V(x) = \begin{cases} 0 & \text{for } |x| < a \\ V_0 > 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

Since V is an even function of x, the Hamiltonian (2.51) commutes with the parity operator P introduced in §4.1.4. So there is a complete set of energy eigenstates of well defined parity. The wavefunctions $u(x) \equiv \langle x|E\rangle$ of these states will be either even or odd functions of x, and this fact will greatly simplify the job of determining $u(x)$.

In the position representation, the governing equation (the TISE 2.33) reads

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2u}{\mathrm{d}x^2} + V(x)u = Eu. \tag{5.2}$$

On account of the step-like nature of V, equation (5.2) reduces to a pair of extremely simple equations,

$$\begin{aligned} \frac{\mathrm{d}^2u}{\mathrm{d}x^2} &= -\frac{2mE}{\hbar^2}u && \text{for } |x| < a \\ \frac{\mathrm{d}^2u}{\mathrm{d}x^2} &= \frac{2m(V_0 - E)}{\hbar^2}u && \text{otherwise.} \end{aligned} \tag{5.3}$$

We restrict ourselves to solutions that describe a particle that is bound by the potential well in the sense that $E < V_0$.[1] Then the solution to the second

---

[1] By considering the behaviour of u near the origin we can prove that $E > 0$.

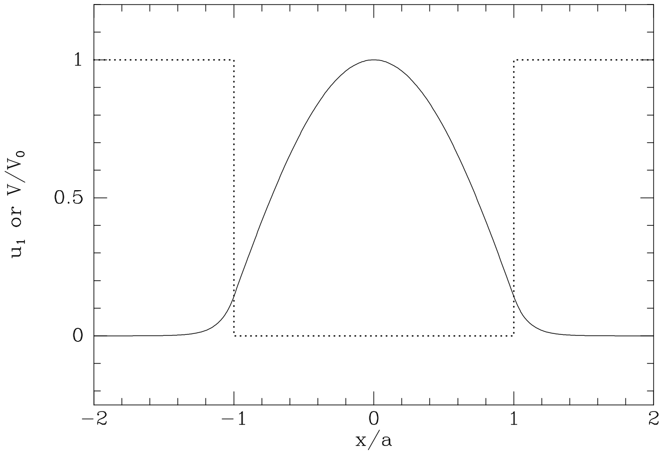


Figure 5.1 The dotted line shows the square-well potential $V(x)$. The full curve shows the ground-state wavefunction.

equation is $u(x) = A\mathrm{e}^{\pm Kx}$, where A is a constant and

$$K \equiv \sqrt{\frac{2m(V_0 - E)}{\hbar^2}}. \tag{5.4}$$

If u is to be normalisable, it must vanish as $|x| \to \infty$. So at $x > 0$ we have $u(x) = A\mathrm{e}^{-Kx}$, and at $x < 0$ we have $u(x) = \pm A\mathrm{e}^{+Kx}$, where the plus sign is required for solutions of even parity, and the minus sign is required for odd parity.

For $E > 0$, the solution to the first of equations (5.3) is either $u(x) = B\cos(kx)$ or $u(x) = B\sin(kx)$ depending on the parity, where

$$k \equiv \sqrt{\frac{2mE}{\hbar^2}}. \tag{5.5}$$

So far we have ensured that $u(x)$ solves the TISE everywhere except at $|x| = a$. Unless u is continuous at these points, $\mathrm{d}u/\mathrm{d}x$ will be arbitrarily large, and $\mathrm{d}^2u/\mathrm{d}x^2$ will be undefined, so u will not satisfy the TISE. Similarly, unless $\mathrm{d}u/\mathrm{d}x$ is continuous at these points, $\mathrm{d}^2u/\mathrm{d}x^2$ will be arbitrarily large, so u cannot solve the TISE. Therefore, we require that both u and $\mathrm{d}u/\mathrm{d}x$ are continuous at $x = a$, that is

$$\left.\begin{aligned} B\cos(ka) &= A\mathrm{e}^{-Ka} \\ -kB\sin(ka) &= -KA\mathrm{e}^{-Ka} \end{aligned}\right\} \quad \text{or} \quad \left\{\begin{aligned} B\sin(ka) &= A\mathrm{e}^{-Ka} \\ kB\cos(ka) &= -KA\mathrm{e}^{-Ka} \end{aligned}\right. \tag{5.6}$$

where the first pair of equations apply in the case of even parity and the second in the case of odd parity. It is easy to show that once these equations have been satisfied, the corresponding equations for $x = -a$ will be automatically satisfied.

We eliminate A and B from equations (5.6) by dividing the second equation in each set by the first. In the case of even parity we obtain

$$k\tan(ka) = K = \sqrt{\frac{2mV_0}{\hbar^2} - k^2}. \tag{5.7}$$

This is an algebraic equation for k, which controls E through (5.5). Before attempting to solve this equation, it is useful to rewrite it as

$$\tan(ka) = \sqrt{\frac{W^2}{(ka)^2} - 1} \quad \text{where} \quad W \equiv \sqrt{\frac{2mV_0a^2}{\hbar^2}}. \tag{5.8}$$

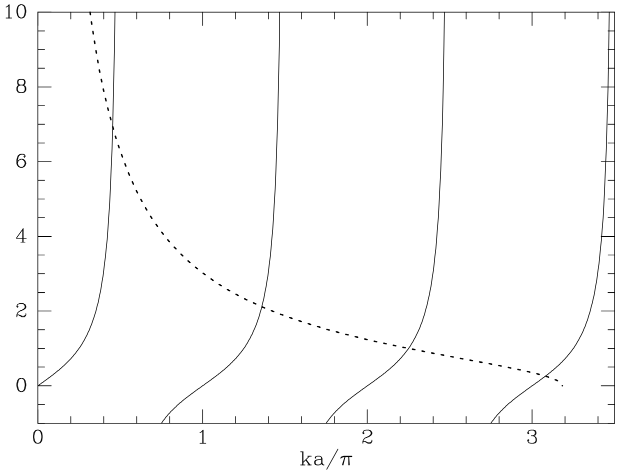


Figure 5.2 Plots of the left (full) and right (dashed) sides of equation (5.8) for the case $W = 10$.

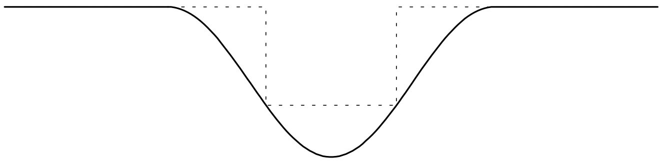


Figure 5.3 A square well inscribed in a general well.

W and ka are dimensionless variables. The left and right sides of equation (5.8) are plotted as functions of ka in Figure 5.2. Since for $ka = 0$ the graphs of the two sides start at the origin and infinity, and the graph of the left side increases to infinity at $ka = \pi/2$ while the graph of the left side terminates at $ka = W$, the equation always has at least one solution. Thus no matter how small V_0 and a are, the square well can always trap the particle. The bigger W is, the more solutions the equation has; a second solution appears at $W = \pi$, a third at $W = 2\pi$, etc.

Analogously one can show that for an odd-parity energy eigenstate to exist, we must have $W > \pi/2$ and that additional solutions appear when $W = (2r + 1)\pi/2$ for $r = 1, 2, \ldots$ (Problem 5.5).

From a naive perspective our discovery that no matter how narrow or shallow it is, a square potential well always has at least one bound state, conflicts with the uncertainty principle: the particle's momentum cannot exceed $p_{\max} = \sqrt{2mE} < \sqrt{2mV_0}$, so if the particle were confined within the well, the product of the uncertainties in p and x would be less than $2ap_{\max} < 2\sqrt{2mV_0a^2} = 2\hbar W$, which tends to zero with W. The resolution of this apparent paradox is that for $W \ll 1$ the particle is not confined *within* the well; there is a non-negligible probability of finding the particle in the classically forbidden region $|x| > a$. In the limit $W \rightarrow 0$ the particle is confined by a well in which it is certain never to be found!

Our result that a square well always has a bound state can be extended to potential wells of any shape: given the potential well U sketched in Figure 5.3, we consider the square well shown by the dashed line in the figure. Since this shallower and narrower well has a bound state, we infer that the potential U also has at least one bound state.

5.1.1 Limiting cases

(a) Infinitely deep well It is worthwhile to investigate the behaviour of these solutions as $V_0 \rightarrow \infty$ with a fixed, when the well becomes infinitely

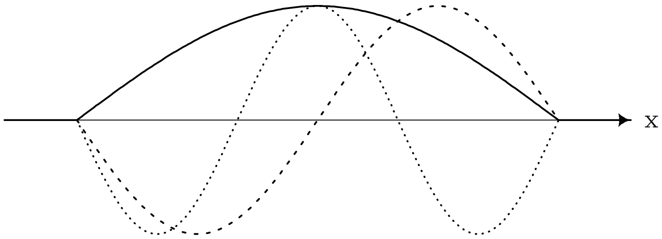


Figure 5.4 The wavefunctions of the lowest three stationary states of the infinitely deep square well: ground state (full); first excited state (dashed); second excited state (dotted).

deep. Then $W \to \infty$ and the dashed curve in Figure 5.2 moves higher and higher up the paper and hits the x axis further and further to the right. Consequently, the values of ka that solve equation (5.8) tend towards $ka = (2r+1)\pi/2$, so the even-parity energy eigenfunctions become

$$u(x) = \begin{cases} A\cos[(2r+1)\pi x/2a] & |x| < a \\ 0 & \text{otherwise.} \end{cases} \tag{5.9}$$

This solution has a discontinuity in its gradient at $x = a$ because it is the limit of solutions in which the curvature K for $x > a$ diverges to infinity. The odd-parity solutions are obtained by replacing the cosine with $\sin(s\pi x/a)$, where $s = 1, 2, \ldots$, which again vanish at the edge of the well (Figure 5.4). From this example we infer the principle that *wavefunctions vanish at the edges of regions of infinite potential energy.*

The energy of any stationary state of an infinite square potential well can be obtained from

$$E_n = \frac{n^2}{8m}\left(\frac{\hbar\pi}{a}\right)^2, \quad \text{where} \quad n = 1, 2, \ldots \tag{5.10}$$

The particle's momentum when it is in the ground state ($n = 1$) is of order $\hbar k = \hbar\pi/2a$ and of undetermined sign, so the uncertainty in the momentum is $\Delta_p \simeq \hbar\pi/a$. The uncertainty in the particle's position is $\Delta_x \simeq 2a$, so $\Delta_x\Delta_p \simeq 2\hbar\pi$, consistent with the uncertainty principle (§2.3.2).

(b) Infinitely narrow well In §11.5.1 we will study a model of covalent bonding that involves the potential obtained by letting the width of the square well tend to zero as the well becomes deeper and deeper in such a way that the product $V_0 a$ remains constant. In this limit $W \propto a\sqrt{V_0}$ (eq. 5.8) tends to zero, so there is only one bound state.

Rather than obtaining the wavefunction and energy of this state from formulae already in hand, it is more convenient to reformulate the problem using a different normalisation for the energy: we now set V to zero outside the well, so V becomes negative at interior points. Then we can write $V(x) = -V_\delta\delta(x)$, where $\delta(x)$ is the Dirac delta function and $V_\delta > 0$. The TISE now reads

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2u}{\mathrm{d}x^2} - V_\delta\delta(x)u = Eu. \tag{5.11}$$

Integrating the equation from $x = -\epsilon$ to $x = \epsilon$ with ϵ infinitesimal, we find

$$-\left[\frac{\mathrm{d}u}{\mathrm{d}x}\right]_{-\epsilon}^{\epsilon} = \frac{2mV_\delta}{\hbar^2}\left[u(0) + E\int_{-\epsilon}^{\epsilon}\mathrm{d}x\,u\right]. \tag{5.12}$$

Since u is finite, the integral on the right can be made as small as we please by letting $\epsilon \to 0$. Hence the content of equation (5.12) is that $\mathrm{d}u/\mathrm{d}x$ has a discontinuity at the origin:

$$\left[\frac{\mathrm{d}u}{\mathrm{d}x}\right]_{-\epsilon}^{\epsilon} = -\frac{2mV_\delta}{\hbar^2}u(0). \tag{5.13}$$

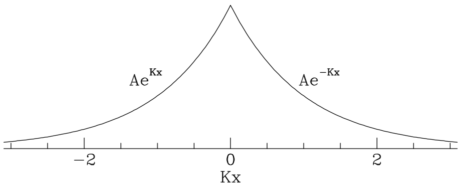


Figure 5.5 Wavefunction of a particle trapped by a very narrow, deep potential well.

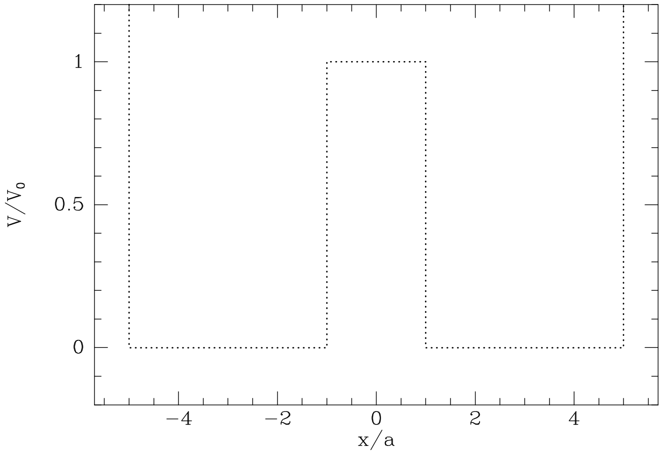


Figure 5.6 A double potential well with $b/a = 5$.

Since we know that the solution we seek has even parity, it is of the form $u(x) = A\mathrm{e}^{\mp Kx}$, where the minus sign applies for $x > 0$ (Figure 5.5). Substituting this form of u into (5.13) and dividing through by $2A$ we have

$$K = \frac{mV_\delta}{\hbar^2}. \tag{5.14}$$

Inserting $u = \mathrm{e}^{-Kx}$ into equation (5.11) at $x > 0$ we find that $E = -\hbar^2K^2/2m$, so the energy of a particle that is bound to a delta-function potential is

$$E = -\frac{mV_\delta^2}{2\hbar^2}. \tag{5.15}$$

5.2 A pair of square wells

Some important phenomena can be illustrated by considering motion in a pair of potentials that are separated by a barrier of finite height and width. Figure 5.6 shows the potential

$$V(x) = \begin{cases} V_0 & \text{for } |x| < a \\ 0 & \text{for } a < |x| < b \\ \infty & \text{otherwise.} \end{cases} \tag{5.16}$$

Since the potential is an even function of x, we may assume that the energy eigenfunctions that we seek are of well-defined parity.

For simplicity we take the potential to be infinite for $|x| > b$, and we assume that the particle is classically forbidden in the region $|x| < a$. Then in this region the wavefunction must be of the form $u(x) = A\cosh(Kx)$ or $u(x) = A\sinh(Kx)$ depending on parity, and K is given by (5.4). In the region $a < x < b$ the wavefunction may be taken to be of the form

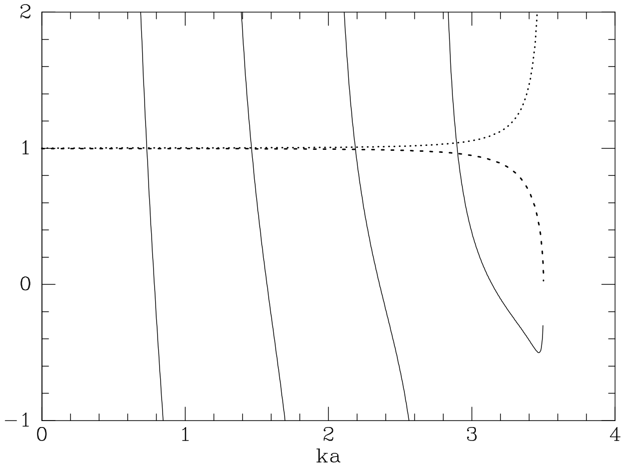


Figure 5.7 Full curves: the left side of equation (5.19) for the case $W = 3.5$, $b = 5a$. Each vertical section is associated with a different value of the integer r. The right side is shown by the dotted curve for even parity, and the dashed curve for odd parity.

$u(x) = B\sin(kx + \phi)$, where B, and ϕ are constants to be determined and k is related to the energy by (5.5). From our study of a single square well we know that u must vanish at $x = b$, so

$$\sin(kb + \phi) = 0 \quad \Rightarrow \quad \phi = r\pi - kb \quad \text{with} \quad r = 0, 1, \dots \tag{5.17}$$

Again by analogy with the case of a single square well, we require u and its derivative to be continuous at $x = a$, so (depending on parity)

$$\left.\begin{aligned} \cosh(Ka) &= B\sin(ka + \phi) \\ K\sinh(Ka) &= kB\cos(ka + \phi) \end{aligned}\right\} \text{ or } \left\{\begin{aligned} \sinh(Ka) &= B\sin(ka + \phi) \\ K\cosh(Ka) &= kB\cos(ka + \phi). \end{aligned}\right. \tag{5.18}$$

Once these equations have been solved, the corresponding conditions at $x = -a$ will be automatically satisfied if for $-b < x < -a$ we take $u = \pm B\sin(k|x| + \phi)$, using the plus sign in the even-parity case.

Using (5.17) to eliminate ϕ from equations (5.18) and then proceeding in close analogy with the working below equations (5.6), we find

$$\tan\left[r\pi - k(b-a)\right]\sqrt{\frac{W^2}{(ka)^2} - 1} = \begin{cases} \coth\left(\sqrt{W^2 - (ka)^2}\right) & \text{even parity} \\ \tanh\left(\sqrt{W^2 - (ka)^2}\right) & \text{odd parity,} \end{cases} \tag{5.19}$$

where W is defined by (5.8).

The left and right sides of equation (5.19) are plotted in Figure 5.7; the values of ka for stationary states correspond to intersections of the steeply sloping curves of the left side with the initially horizontal curves of the right side. The smallest value of ka is associated with the ground state. The values come in pairs, one for an even-parity state, and one very slightly larger for an odd-parity state. The difference between the k values in a pair increases with k.

The closeness of the k values in a given pair ensures that in the right-hand well ($a < x < b$) the wavefunctions of the even- and odd-parity states are very similar, and that in the left-hand well the wavefunctions differ by little more than sign – see Figure 5.8. Moreover, when the k values are similar, the amplitude of the wavefunction is small in the classically forbidden region $|x| < a$. Hence, the linear combinations

$$\psi_\pm(x) \equiv \frac{1}{\sqrt{2}}\left[u_\mathrm{e}(x) \pm u_\mathrm{o}(x)\right] \tag{5.20}$$

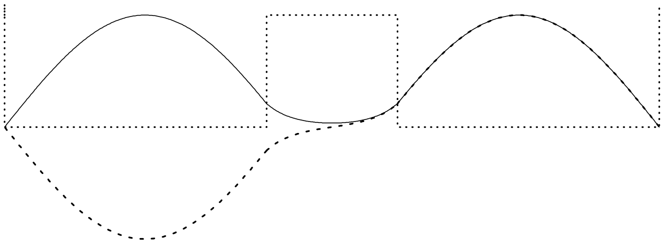


Figure 5.8 The ground state (full curve) and the associated odd-parity state (dashed curve) of the double square-well potential (shown dotted).

where $u_e(x)$ is the even-parity wavefunction of a pair, describe states ψ_+ in which the particle is almost certain to be in the right-hand well, and ψ_- in which it is equally certain to be in the left-hand well.

Consider now how the system evolves if at time 0 it is in the state ψ_+, so the particle is in the right-hand well. Then by equation (2.32) at time t its wavefunction is

$$\begin{aligned}\psi(t) &= \frac{1}{\sqrt{2}}\left[u_e(x)\mathrm{e}^{-\mathrm{i}E_e t/\hbar} + u_o(x)\mathrm{e}^{-\mathrm{i}E_o t/\hbar}\right]\\ &= \frac{1}{\sqrt{2}}\mathrm{e}^{-\mathrm{i}E_e t/\hbar}\left[u_e(x) + u_o(x)\mathrm{e}^{-\mathrm{i}(E_o-E_e)t/\hbar}\right].\end{aligned} \tag{5.21}$$

After a time $T = \pi\hbar/(E_o - E_e)$ the exponential in the square brackets on the second line of this equation equals -1, so to within an overall phase factor the wavefunction has become $[u_e(x) - u_o(x)]/\sqrt{2}$, implying that the particle is certainly in the left-hand well; we say that in the interval T the particle has **tunnelled** through the barrier that divides the wells. After a further period T it is certainly in the right-hand well, and so on ad infinitum. In classical physics the particle would stay in whatever well it was in initially. In fact, the position of a familiar light switch is governed by a potential that consists of two similar adjacent potential wells, and such switches most definitely do not oscillate between their on and off positions. We do not observe tunnelling in the classical regime because $E_o - E_e$ decreases with increasing W faster than e^{-2W} (Problem 5.13), so the time required for tunnelling to occur increases faster than e^{2W} and is enormously long for classical systems such as light switches.

5.2.1 Ammonia

Nature provides us with a beautiful physical realisation of a system with a double potential well in the ammonia molecule NH_3. Ammonia contains four nuclei and ten electrons, so is really a very complicated dynamical system. However, in §11.5.2 we shall show that a useful way of thinking about the low-energy behaviour of molecules is to imagine that the electrons provide light springs, which hold the nuclei together. The nuclei oscillate around the equilibrium positions defined by the potential energy of these springs. In the case of NH_3, the potential energy is minimised when the three hydrogen atoms are arranged at the vertices of an equilateral triangle, while the nitrogen atom lies some distance x away from the plane of the triangle, either 'above' or 'below' it (see Figure 5.9). Hence if we were to plot the molecule's potential energy as a function of x, we would obtain a graph that looked like Figure 5.6 except that the sides of the wells would be sloping rather than straight. This function would yield eigenenergies that came in pairs, as in our square-well example.

In many physical situations the molecule would have so little energy that it could have negligible amplitudes to be found in any but the two lowest-lying stationary states, and we would obtain an excellent approximation to

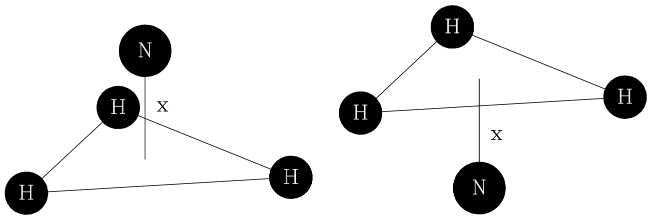


Figure 5.9 The two possible relative locations of nitrogen and hydrogen atoms in NH_3.

the dynamics of ammonia by including only the amplitudes to be found in these two states. We now use Dirac notation to study this dynamics.

Let $|+\rangle$ be the state whose wavefunction is analogous to the wavefunction $\psi_+(x)$ defined above in the case of the double square well; then $\psi_+(x) = \langle x|+\rangle$, and in the state $|+\rangle$ the N atom is certainly above the plane containing the H atoms. The ket $|-\rangle$ is the complementary state in which the N atom lies below the plane of the H atoms.

The $|\pm\rangle$ states are linear combinations of the eigenkets $|\mathrm{e}\rangle$ and $|\mathrm{o}\rangle$ of the Hamiltonian:

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|\mathrm{e}\rangle \pm |\mathrm{o}\rangle). \tag{5.22}$$

In the $|\pm\rangle$ basis the matrix elements of the Hamiltonian H are

$$\begin{aligned}\langle +|H|+\rangle &= \tfrac{1}{2}(\langle \mathrm{e}| + \langle \mathrm{o}|)H(|\mathrm{e}\rangle + |\mathrm{o}\rangle) = \tfrac{1}{2}(E_\mathrm{e} + E_\mathrm{o})\\ \langle +|H|-\rangle &= \tfrac{1}{2}(\langle \mathrm{e}| + \langle \mathrm{o}|)H(|\mathrm{e}\rangle - |\mathrm{o}\rangle) = \tfrac{1}{2}(E_\mathrm{e} - E_\mathrm{o})\\ \langle -|H|-\rangle &= \tfrac{1}{2}(\langle \mathrm{e}| - \langle \mathrm{o}|)H(|\mathrm{e}\rangle - |\mathrm{o}\rangle) = \tfrac{1}{2}(E_\mathrm{e} + E_\mathrm{o})\end{aligned} \tag{5.23}$$

Bearing in mind that H is represented by a Hermitian matrix, we conclude that it is

$$H = \begin{pmatrix} \overline{E} & -A \\ -A & \overline{E} \end{pmatrix}, \tag{5.24}$$

where $\overline{E} = \frac{1}{2}(E_\mathrm{e} + E_\mathrm{o})$ and $A \equiv \frac{1}{2}(E_\mathrm{o} - E_\mathrm{e})$ are both positive.

Now the electronic structure of NH_3 is such that the N atom carries a small negative charge $-q$, with a corresponding positive charge $+q$ distributed among the H atoms. With NH_3 in either the $|+\rangle$ or $|-\rangle$ state there is a net separation of charge, so an ammonia molecule in these states possesses an electric dipole moment of magnitude qs directed perpendicular to the plane of H atoms (see Figure 5.9), where s is a small distance.

Below equation (5.21) we saw that a molecule that is initially in the state $|+\rangle$ will subsequently oscillate between this state and the state $|-\rangle$ at a frequency $(E_\mathrm{o} - E_\mathrm{e})/2\pi\hbar = A/\pi\hbar$. Hence a molecule that starts in the state $|+\rangle$ is an oscillating dipole and it will emit electromagnetic radiation at the frequency $A/\pi\hbar$. This proves to be 150 GHz, so the molecule emits microwave radiation.

The ammonia maser The energy $2A$ that separates the ground and first excited states of ammonia in zero electric field is small, 10^{-4} eV. Consequently at room temperature similar numbers of molecules are in these two states. The principle of an ammonia maser is to isolate the molecules that are in the first excited state, and then to harvest the radiation that is emitted as the molecules decay to the ground state. The isolation is achieved by exploiting the fact that, as we now show, when an electric field is applied, molecules in the ground and first excited states develop polarisations of opposite sign.

We define the dipole-moment operator P by

$$P|+\rangle = -qs|+\rangle \quad ; \quad P|-\rangle = +qs|-\rangle, \tag{5.25}$$

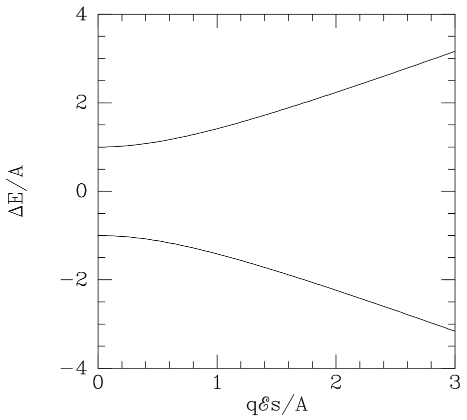


Figure 5.10 Energy levels of the ammonia molecule as a function of external electric field strength $\mathcal{E}$. The quantity plotted, $\Delta E = E - \overline{E}$.

so a molecule in the $|+\rangle$ state has dipole moment $-qs$ and a molecule in the $|-\rangle$ state has dipole moment $+qs$.[2] To measure this dipole moment, we can place the molecule in an electric field of magnitude $\mathcal{E}$ parallel to the dipole axis. Since the energy of interaction between a dipole P and an electric field $\mathcal{E}$ is $-P\mathcal{E}$, the new Hamiltonian is

$$H = \begin{pmatrix} \overline{E} + q\mathcal{E}s & -A \\ -A & \overline{E} - q\mathcal{E}s \end{pmatrix}. \tag{5.26}$$

This new Hamiltonian has eigenvalues

$$E_\pm = \overline{E} \pm \sqrt{A^2 + (q\mathcal{E}s)^2}. \tag{5.27}$$

These are plotted as a function of field $\mathcal{E}$ in Figure 5.10. When $\mathcal{E} = 0$ the energy levels are the same as before. As $\mathcal{E}$ slowly increases, E increases quadratically with $\mathcal{E}$, because $\sqrt{A^2 + (q\mathcal{E}s)^2} \simeq A + (q\mathcal{E}s)^2/2A$, but when $\mathcal{E} \gg A/qs$ the energy eigenvalues change linearly with $\mathcal{E}$. Notice that in this large-field limit, at lowest order the energy levels do not depend on A.

The physical interpretation of these results is the following. In the absence of an electric field, the energy eigenstates are the states of well-defined parity $|\mathrm{e}\rangle$ and $|\mathrm{o}\rangle$, which have no dipole moment. An electric field breaks the symmetry between the two potential wells, making it energetically favourable for the N atom to occupy the well to which the electric field is pushing it. Consequently, the ground state develops a dipole moment P, which is proportional to $\mathcal{E}$. Thus at this stage the electric contribution to the energy of the ground state, which is $-P\mathcal{E}$, is proportional to $\mathcal{E}^2$. Once this contribution exceeds the separation A between the states of well-defined parity, the molecule has shifted to the lower-energy state of the pair $|\pm\rangle$, and it stays in this state as the electric field is increased further. Thus for large fields the polarisation of the ground state is independent of $\mathcal{E}$ and the electric contribution to the energy is simply proportional to $\mathcal{E}$.

While the ground state develops a dipole moment that lowers its energy, the first excited state develops the opposite polarisation, so the electric field raises the energy of this state, as shown in Figure 5.10. The response of the first excited state is anomalous from a classical perspective.

Ehrenfest's theorem (2.57) tells us that the expectation values of operators obey classical equations of motion. In particular the momentum of a molecule obeys

$$\frac{\mathrm{d}\langle p_x\rangle}{\mathrm{d}t} = -\left\langle \frac{\partial V}{\partial x} \right\rangle, \tag{5.28}$$

---

[2] The N atom is negatively charged so the dipole points away from it.

where x is a Cartesian coordinate of the molecule's centre of mass. The potential depends on x only through the electric field $\mathcal{E}$, so

$$\frac{\partial V}{\partial x} = -P\frac{\partial \mathcal{E}}{\partial x}, \tag{5.29}$$

from which it follows that

$$\frac{\mathrm{d}\langle p_x\rangle}{\mathrm{d}t} = \langle P\rangle \frac{\partial \mathcal{E}}{\partial x}. \tag{5.30}$$

Since the sign of $\langle P\rangle$ and therefore the force on a molecule depends on whether the molecule is in the ground or first excited state, when a jet of ammonia passes through a region of varying $\mathcal{E}$, molecules in the first excited state can be separated from those in the ground state.

Having gathered the molecules that are in the excited state, we lead them to a cavity that resonates with the 150 GHz radiation that is emitted when molecules drop into the ground state. The operation of an ammonia maser by Charles Townes and colleagues[3] was the first demonstration of stimulated emission and opened up the revolution in science and technology that lasers have have since wrought.

5.3 Scattering of free particles

We now consider what happens when particles that are travelling parallel to the x axis encounter a sudden change in the potential. We consider two cases: in one case the particles encounter a square potential well such as we discussed in §5.1, and in the other case they encounter a potential barrier. In classical physics all the particles pass through the potential well in the first case, while all the particles are reflected in the second case. We shall find that quantum mechanics predicts that in each case some particles are reflected and some are transmitted.

We consider particles that have well-defined energy E, so their wavefunction satisfies the TISE (5.2).

5.3.1 Reflection off a potential well

We take the potential to be (Figure 5.11)

$$V = \begin{cases} -V_0 & \text{for } |x| < a \\ 0 & \text{otherwise,} \end{cases} \tag{5.31}$$

where $V_0 > 0$. For this potential the relevant solutions of (5.2) are

$$u(x) = \begin{cases} A_+\mathrm{e}^{\mathrm{i}kx} + A_-\mathrm{e}^{-\mathrm{i}kx} & \text{for } x < a \\ B_+\mathrm{e}^{\mathrm{i}Kx} + B_-\mathrm{e}^{-\mathrm{i}Kx} & \text{for } |x| < a \\ C\mathrm{e}^{\mathrm{i}kx} & \text{for } x > a, \end{cases} \tag{5.32a}$$

where

$$k^2 = \frac{2mE}{\hbar^2} \quad ; \quad K^2 = \frac{2m(E+V_0)}{\hbar^2}. \tag{5.32b}$$

Since the wavefunction of a particle with well-defined momentum $p = \hbar k$ is $\mathrm{e}^{\mathrm{i}kx}$ (eq. 2.61), in the first line on the right of (5.32a), A_+ is proportional to the amplitude for the particle to be moving towards the barrier. Hence A_+ is a measure of the flux of incoming particles, while A_- is a measure of the flux of particles that have been reflected by the barrier. Similarly, in the

---

[3] Gordon, J.P., Zeiger, H.J., & Townes, C.H., 1954, Phys. Rev, **95**, 282 (1954)

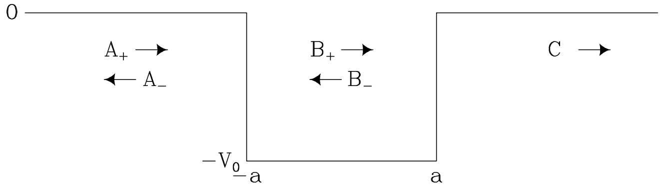


Figure 5.11 A square potential well encountered by particles incident from the left.

middle line $B_\pm$ are measures of the numbers of particles that are moving in each direction within the well, and in the third line C is a measure of the flux of particles that have passed through the well.

As in §5.1, we require u to be continuous and to have continuous derivatives at both edges of the well. Thus

$$\begin{aligned} A_+\mathrm{e}^{-\mathrm{i}ka}+A_-\mathrm{e}^{\mathrm{i}ka} &= B_+\mathrm{e}^{-\mathrm{i}Ka}+B_-\mathrm{e}^{\mathrm{i}Ka}\\ k(A_+\mathrm{e}^{-\mathrm{i}ka}-A_-\mathrm{e}^{\mathrm{i}ka}) &= K(B_+\mathrm{e}^{-\mathrm{i}Ka}-B_-\mathrm{e}^{\mathrm{i}Ka})\\ C\mathrm{e}^{\mathrm{i}ka} &= B_+\mathrm{e}^{\mathrm{i}Ka}+B_-\mathrm{e}^{-\mathrm{i}Ka}\\ kC\mathrm{e}^{\mathrm{i}ka} &= K(B_+\mathrm{e}^{\mathrm{i}Ka}-B_-\mathrm{e}^{-\mathrm{i}Ka}). \end{aligned}\tag{5.33}$$

These four equations between the five unknowns in the problem enable us to express four of the unknowns as functions of the fifth, say A_+. This mathematical situation corresponds to the physical reality that the flux of incoming particles can be controlled, but then all the other fluxes are determined by the physics of the problem.

A reasonable strategy for solving the equations is to eliminate C by dividing the fourth equation into the third, thus establishing a relationship between B_+ and B_-:

$$B_-=\frac{K/k-1}{K/k+1}\mathrm{e}^{2\mathrm{i}Ka}B_+.\tag{5.34}$$

Then we substitute this relation into the first two equations to obtain

$$\begin{aligned} A_+\mathrm{e}^{-\mathrm{i}ka}+A_-\mathrm{e}^{\mathrm{i}ka} &= B_+\left(\mathrm{e}^{-\mathrm{i}KA}+\frac{K/k-1}{K/k+1}\mathrm{e}^{3\mathrm{i}Ka}\right)\\ A_+\mathrm{e}^{-\mathrm{i}ka}-A_-\mathrm{e}^{\mathrm{i}ka} &= B_+\left(\mathrm{e}^{-\mathrm{i}KA}-\frac{K/k-1}{K/k+1}\mathrm{e}^{3\mathrm{i}Ka}\right)\frac{K}{k}. \end{aligned}\tag{5.35}$$

Now we divide the first of these equations by the second to obtain a relation between A_- and A_+:

$$A_-=A_+\,\mathrm{e}^{-2\mathrm{i}ka}\,\frac{Q-1}{Q+1},\tag{5.36a}$$

where

$$Q\equiv\frac{\cos 2Ka-\mathrm{i}(k/K)\sin 2Ka}{\cos 2Ka-\mathrm{i}(K/k)\sin 2Ka}.\tag{5.36b}$$

The condition for A_- to vanish, as it would in the classical case, is $Q=1$. This condition is satisfied when $K=k$ because $V_0=0$ and there is no step in the potential, or $2Ka=r\pi$ for $r=0,1,\ldots$. In general these conditions will not be met and in quantum mechanics we expect some particles to be reflected by the well, contrary to our classical expectations. Figure 5.12 plots the reflection probability $|A_-/A_+|^2$ as a function of E from equations (5.36); it is made up of oscillations superposed on a general decline with increasing E. The oscillations are caused by quantum interference between the amplitudes that are reflected by the two discontinuities in $V(x)$; the

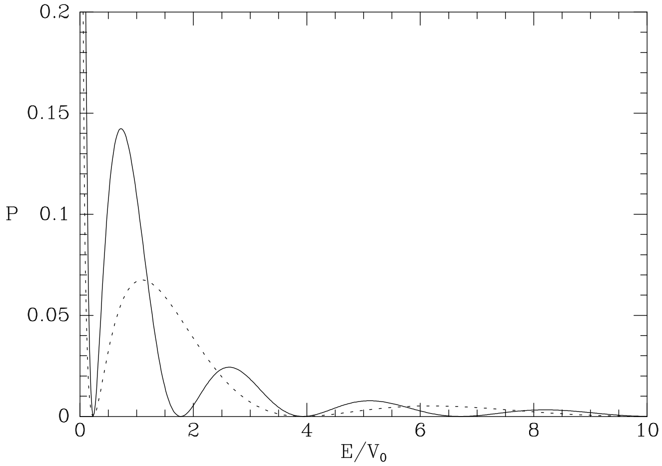


Figure 5.12 The probability of reflection by a square potential well of depth V_0 and width a (eqns 5.36). The full curve is for $a\sqrt{2mV_0}/\hbar = 2$, while the dashed curve is for $a\sqrt{2mV_0}/\hbar = 1$.

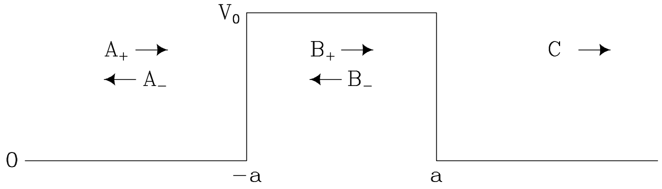


Figure 5.13 A square barrier encountered by particles incident from the left.

general decline arises because the amplitudes of the reflected waves decrease as E/V_0 grows.

When we eliminate A_- from (5.35), we obtain

$$\frac{B_+}{A_+} = \frac{Q}{Q+1}\left(1+\frac{k}{K}\right)\frac{e^{-i(k+K)a}}{\cos 2Ka - i(k/K)\sin 2Ka}. \tag{5.37}$$

When $2Ka = r\pi$ so no particles are reflected, this relation yields

$$\left|\frac{B_+}{A_+}\right| = \tfrac{1}{2}\left(1+\frac{k}{K}\right). \tag{5.38}$$

Consider the case $K \gg k$ of a very deep well. Then $|B_+/A_+| \to \frac{1}{2}$. The square of this number is the ratio of the density of particles that are moving to the right in the potential well to the density of incident particles. Since the speed of the particles is $\hbar k/m$ at $x < -a$ and $\hbar K/m$ in the well, the flux of particles that are moving to the right in the well is larger than the flux of incident particles by a factor $\frac{1}{4}K/k \gg 1$. Hence, even though no particles are reflected by the well, few particles pass straight through it: when the well is deep a particle typically rattles to and fro through the well many times before escaping to $x > a$.

5.3.2 Tunnelling through a potential barrier

We now consider what happens when a stream of particles of well defined energy E impacts from $x < 0$ on a barrier. We take the potential to be (Figure 5.13)

$$V = \begin{cases} V_0 & \text{for } |x| < a \\ 0 & \text{otherwise,} \end{cases} \tag{5.39}$$

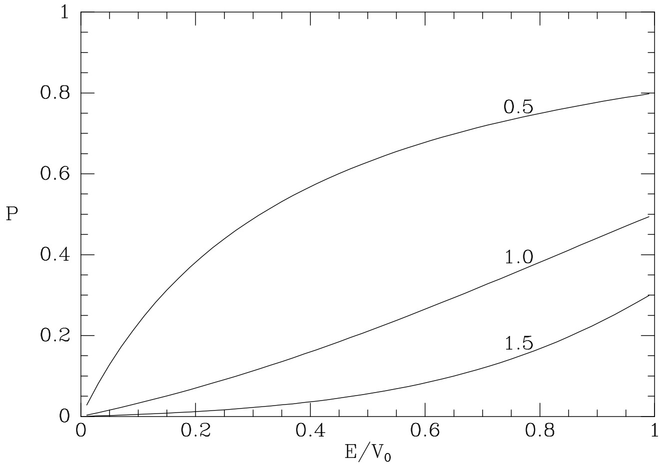


Figure 5.14 The probability of passing a barrier of height V_0 and width a as a function of energy. The curves are labelled by the values of $\sqrt{2mV_0}a/\hbar$.

where $V_0 > E$. For this potential the relevant solutions of (5.2) are

$$u(x) = \begin{cases} A_+ \mathrm{e}^{\mathrm{i}kx} + A_- \mathrm{e}^{-\mathrm{i}kx} & \text{for } x < a \\ B_+ \mathrm{e}^{Kx} + B_- \mathrm{e}^{-Kx} & \text{for } |x| < a \\ C\mathrm{e}^{\mathrm{i}kx} & \text{for } x > a, \end{cases} \tag{5.40a}$$

where

$$k^2 = \frac{2mE}{\hbar^2} \quad ; \quad K^2 = \frac{2m(V_0 - E)}{\hbar^2}. \tag{5.40b}$$

This trial wavefunction differs from that (5.32a) used in the last subsection only in the region $|x| < a$, where the exponentials now have real arguments. The analysis proceeds exactly as before: we require u to be continuous and have a continuous derivative at $x = \pm a$. From the conditions at $x = a$, we obtain

$$B_- = -\frac{1 - K/\mathrm{i}k}{1 + K/\mathrm{i}k} B_+. \tag{5.41}$$

Using this relation to eliminate B_- from the conditions for $x = -a$, and then eliminating B_+ between the conditions yields

$$A_- = A_+ \mathrm{e}^{-2\mathrm{i}ka} \frac{\widetilde{Q} - 1}{\widetilde{Q} + 1}, \tag{5.42a}$$

where

$$\widetilde{Q} \equiv \frac{\cosh 2Ka - (\mathrm{i}k/K)\sinh 2Ka}{\cosh 2Ka - (K/\mathrm{i}k)\sinh 2Ka}. \tag{5.42b}$$

$\widetilde{Q}$ is never unity, so some particles are always reflected.

By conservation of particles (which may be explicitly checked by calculating C and proving that $|A_+|^2 - |A_-|^2 = |C|^2$), the fraction of particles that get through the barrier is

$$\begin{aligned} \frac{|A_+|^2 - |A_-|^2}{|A_+|^2} &= 1 - \left|\frac{\widetilde{Q} - 1}{\widetilde{Q} + 1}\right|^2 \\ &= \frac{1}{\cosh^2 2Ka + \frac{1}{4}(k/K - K/k)^2 \sinh 2Ka}. \end{aligned} \tag{5.43}$$

This fraction is plotted in Figure 5.14 for three values of the dimensionless parameter $\sqrt{2mV_0}a/\hbar$. At small values of E/V_0, $Ka \gg 1$ and the denominator on the right of equation (5.43) is dominated by $\cosh^2(2Ka) \sim \frac{1}{4}\mathrm{e}^{4Ka}$, so the probability of penetration decreases very rapidly as a increases. As E approaches V_0, which in classical physics is sufficient for the particles to get over the barrier, the quantum penetration probability rises to a value that is less than unity, even for the narrowest barrier because the sudden increase in the potential at the left-hand edge of the barrier reflects some of the incoming particles even when $E > V_0$ (Problem 5.7).

5.4 What we have learnt

In this chapter we have examined some highly idealised systems. Let's take a moment so summarise what we have learnt from them, and to ask how widely these results can be applied. We have found the following:

- A square potential well has at least one bound state, and may have more depending on the size of the value of the dimensionless parameter W (eq. 5.8).
- Any potential well has at least one bound state.
- A particle trapped by a very narrow or shallow well has negligible probability of being found in the well.
- When solving the TISE in the presence of an infinite step in the potential, we should require the wavefunction to vanish at the base of the step.
- When two identical square potential wells are separated by a barrier, the eigenenergies occur in pairs, and the associated wavefunctions have either even or odd parity with respect to an origin that is symmetrically placed between the wells. The even-parity state lies slightly lower in energy than the odd-parity state. A sum of the lowest two eigenstates is a state in which the particle is certainly in one well, while the difference gives a state in which the particle is certainly in the other well. A particle that starts in one well oscillates between the wells with a period inversely proportional to the energy difference between the energy eigenstates. The particle is said to 'tunnel' through the barrier that divides the two wells at a rate that decreases exponentially with the product of the barrier's height and the width.
- In an ammonia molecule the nitrogen atom moves in an effective potential that provides two identical wells and the above model explains how an ammonia maser works.
- When particles encounter a sharp drop in the potential, there is a non-negligible probability that they will be reflected, even though classical physics predicts that none would be reflected. When there are two sudden changes in the potential, the overall reflection probability oscillates on account of quantum interference between the amplitudes to be reflected at the two discontinuities.
- When particles encounter a potential barrier, they have a non-zero probability to cross this barrier, even when it is too high for them to be able to cross according to classical physics. Similarly, when the barrier is low enough for a classical crossing, they generally have a non-zero probability of being reflected by the barrier.

We must now ask "how important for these results is the assumption that the potential changes discontinuously?" Indeed, it seems unlikely that any real system has a discontinuous potential $V(x)$, so our results are of practical interest only if sufficiently steep changes in V can be treated as discontinuous.

When the wavefunction is **evanescent** (i.e. $\propto \mathrm{e}^{\pm x}$) on one side of the discontinuity, our results carry over to potentials that change continuously: where $E < V(x)$, the wavefunction is no longer a simple exponential but its phase remains constant and its amplitude decreases monotonically, while in the region $E > V(x)$ the sinusoidal dependence $\psi(x) \propto \mathrm{e}^{\mathrm{i}kx}$ is replaced by some other oscillatory function of similar amplitude. Qualitatively the

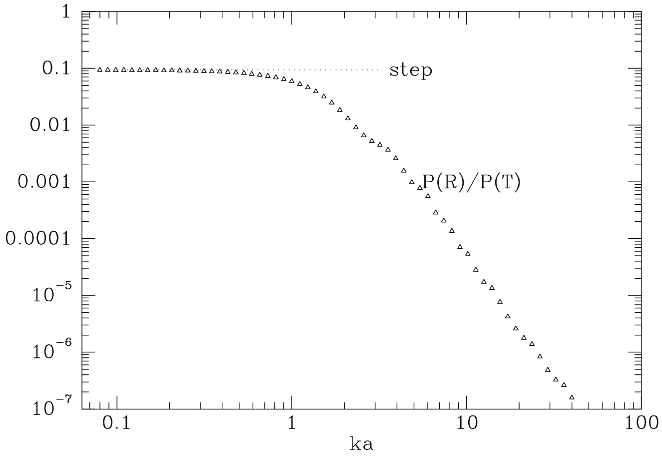


Figure 5.15 The symbols show the ratio of the probability of reflection to the probability of transmission when particles move from $x = -\infty$ in the potential (5.44) with energy $E = \hbar^2 k^2/2m$ and $V_0 = 0.7E$. The dotted line is the value obtained for a step change in the potential (Problem 5.4).

results we have obtained for particles confined by a step potential carry over to continuously varying potentials when $E < V$ on one side of a region of varying potential.

The situation is less clear when $E > V(x)$ on both sides of the change in the potential. The relevance to such cases of our solutions for step potentials can be investigated by solving the TISE numerically for a potential that changes over a distance that can be varied. Consider for example

$$V(x) = V_0 \begin{cases} 0 & \text{for } x < -a \\ \frac{1}{2}[1 + \sin(\pi x/2a)] & \text{for } |x| < a \\ 1 & \text{for } x > a, \end{cases} \tag{5.44}$$

which changes from 0 to V_0 over a distance $2a$ centred on the origin. Figure 5.15 shows that when particles with energy $\hbar^2 k^2/2m > V_0$ are incident on this step from $x = -\infty$, the probability that a particle is reflected is very small for $ka \gtrsim 2$.[4] For $ka \ll 1$, the probability of reflection is close to that obtained for the corresponding step potential (Problem 5.4). We conclude that *results obtained for step potentials can be applied to continuous potentials so long as the distance over which the potential changes is small compared to the de Broglie wavelength* $\lambda = h/p$ *of the incident particles.*

Let's consider under what circumstances this condition could be satisfied for a stream of electrons. The de Broglie wavelength of electrons with kinetic energy E is

$$\lambda = 1.16 \times 10^{-9} \left(\frac{E}{1\,\mathrm{eV}}\right)^{-1/2} \mathrm{m}. \tag{5.45}$$

For the one-dimensional approximation to apply, we need the beam to be many λ wide, and for the step approximation to be valid, we require the change in potential to be complete well inside λ. In practice these conditions can be simultaneously satisfied only when the potential change is associated with a change in the medium through which the electrons are propagating. If the medium is made of atoms, the change must extend over at least the characteristic size of atoms 0.1 nm. Hence we require $E \lesssim 1\,\mathrm{eV}$.

Realistically step potentials are relevant only for less massive particles, photons and neutrinos. The propensity for some photons to be transmitted and some reflected at an abrupt change in potential, such as that at a

---

[4] The 'WKBJ' approximation derived in §11.6 provides an analytic approximation to the solution of the TISE when ka is significantly larger than 2π.

glass/air interface, plays an important role in optics. By contrast, electrons, neutrons and protons are unlikely to be partially transmitted and partially reflected by a region of varying potential.

These considerations explain why the phenomenon of partial reflection and partial transmission is unknown to classical mechanics, which is concerned with massive bodies that have de Broglie wavelengths many orders of magnitude smaller than an atom at any experimentally accessible energy.

Problems

5.1 A particle is confined by the potential well

$$V(x) = \begin{cases} 0 & \text{for } |x| < a \\ \infty & \text{otherwise.} \end{cases} \tag{5.46}$$

Explain (a) why we can assume that there is a complete set of stationary states with well-defined parity and (b) why to find the stationary states we solve the TISE subject to the boundary condition $\psi(\pm a) = 0$.

Determine the particle's energy spectrum and give the wavefunctions of the first two stationary states.

5.2 At $t = 0$ the particle of Problem 5.1 has the wavefunction

$$\psi(x) = \begin{cases} 1/\sqrt{2a} & \text{for } |x| < a \\ 0 & \text{otherwise.} \end{cases} \tag{5.47}$$

Find the probabilities that a measurement of its energy will yield: (a) $9\hbar^2\pi^2/(8ma^2)$; (b) $16\hbar^2\pi^2/(8ma^2)$.

5.3 Find the probability distribution of measuring momentum p for the particle described in Problem 5.2. Sketch and comment on your distribution. Hint: express $\langle p|x\rangle$ in the position representation.

5.4 Particles move in the potential

$$V(x) = \begin{cases} 0 & \text{for } x < 0 \\ V_0 & \text{for } x > 0. \end{cases} \tag{5.48}$$

Particles of mass m and energy $E > V_0$ are incident from $x = -\infty$. Show that the probability that a particle is reflected is

$$\left(\frac{k-K}{k+K}\right)^2, \tag{5.49}$$

where $k \equiv \sqrt{2mE}/\hbar$ and $K \equiv \sqrt{2m(E-V_0)}/\hbar$. Show directly from the TISE that the probability of transmission is

$$\frac{4kK}{(k+K)^2} \tag{5.50}$$

and check that the flux of particles moving away from the origin is equal to the incident particle flux.

5.5 Show that the energies of odd-parity energy eigenfunctions of the square potential well

$$V(x) = \begin{cases} 0 & \text{for } |x| < a \\ V_0 > 0 & \text{otherwise,} \end{cases} \tag{5.51}$$

are governed by

$$\cot(ka) = -\sqrt{\frac{W^2}{(ka)^2} - 1} \quad \text{where} \quad W \equiv \sqrt{\frac{2mV_0a^2}{\hbar^2}} \quad \text{and} \quad k^2 = 2mE/\hbar^2. \tag{5.52}$$

Show that for an odd-parity state to exist, we require $W > \pi/2$.

5.6 Give an example of a potential in which there is a complete set of stationary states of well-defined parity, and an alternative complete set of stationary states that are not eigenkets of the parity operator. Hint: think about NH_3.

5.7 Obtain the probability of passing through the square barrier

$$V = \begin{cases} V_0 & \text{for } |x| < a \\ 0 & \text{otherwise,} \end{cases} \tag{5.53}$$

in the case $E > V_0 > 0$. Verify that the probability you obtain joins at $E = V_0$ to the transition probability for $E < V_0$, which is

$$P(E) = \frac{1}{\cosh^2 2Ka + \frac{1}{4}(k/K - K/k)^2 \sinh 2Ka} \quad \text{with} \quad K^2 \equiv \frac{2m(V_0 - E)}{\hbar^2}. \tag{5.54}$$

5.8 In this problem you find stationary states of well-defined parity for a free particle that moves in the potential

$$V = \begin{cases} -V_0 & \text{for } |x| < a \\ 0 & \text{otherwise} \end{cases} \quad (V_0 > 0). \tag{5.55}$$

For $|x| < a$ the (unnormalised) wavefunctions are clearly $u_{\rm e}(x) = \cos(Kx)$ and $u_{\rm o} = \sin(Kx)$. For $x > a$ write $u_{\rm e}(x) = A\cos(kx + \phi_{\rm e})$ and $u_{\rm o}(x) = B\sin(kx + \phi_{\rm o})$. Write down expressions for K and k in terms of the energy E. Write down the matching conditions at $x = a$ and explain why in contrast to the bound case these can be solved for any energy E.

A general state of energy E can be written $\psi(x) = Cu_{\rm e}(x) + Du_{\rm o}(x)$. Show that the condition for ψ to have only outgoing probability at $x > a$ is $DB/CA = {\rm i}{\rm e}^{{\rm i}(\phi_{\rm o} - \phi_{\rm e})}$, and that the coefficient of ${\rm e}^{{\rm i}kx}$ in $\psi(x)$ is then

$$\tfrac{1}{2}CA{\rm e}^{{\rm i}\phi_{\rm e}}\left(1 + {\rm e}^{2{\rm i}(\phi_{\rm o} - \phi_{\rm e})}\right).$$

Write down the forms of $u_{\rm e}(x)$ and $u_{\rm o}(x)$ that are valid at $x < -a$ and hence show that at $x < -a$ the coefficient of ${\rm e}^{{\rm i}kx}$ in $\psi(x)$ is $CA{\rm e}^{-{\rm i}\phi_{\rm e}}$. Hence give an expression for the transmission probability and show that the particle is certain to cross the well if $Ka = r\pi$.

5.9 An electron moves along an infinite chain of potential wells. For sufficiently low energies we can assume that the set $\{|n\rangle\}$ is complete, where $|n\rangle$ is the state of definitely being in the $n^{\rm th}$ well. By analogy with our analysis of the NH_3 molecule we assume that for all n the only non-vanishing matrix elements of the Hamiltonian are $\mathcal{E} \equiv \langle n|H|n\rangle$ and $A \equiv \langle n \pm 1|H|n\rangle$. Give physical interpretations of the numbers A and $\mathcal{E}$.

Explain why we can write

$$H = \sum_{n=-\infty}^{\infty} \mathcal{E}|n\rangle\langle n| + A\left(|n\rangle\langle n+1| + |n+1\rangle\langle n|\right). \tag{5.56}$$

Writing an energy eigenket $|E\rangle = \sum_n a_n|n\rangle$ show that

$$a_m(E - \mathcal{E}) - A\left(a_{m+1} + a_{m-1}\right) = 0. \tag{5.57}$$

Obtain solutions of these equations in which $a_m \propto {\rm e}^{{\rm i}km}$ and thus find the corresponding energies E_k. Why is there an upper limit on the values of k that need be considered?

Initially the electron is in the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|E_k\rangle + |E_{k+\Delta}\rangle\right), \tag{5.58}$$

where $0 < k \ll 1$ and $0 < \Delta \ll k$. Describe the electron's subsequent motion in as much detail as you can.

5.10* In this problem you investigate the interaction of ammonia molecules with electromagnetic waves in an ammonia maser. Let $|+\rangle$ be the state in which the N atom lies above the plane of the H atoms and $|-\rangle$ be the state in which the N lies below the plane. Then when there is an oscillating electric field $\mathcal{E}\cos\omega t$ directed perpendicular to the plane of the hydrogen atoms, the Hamiltonian in the $|\pm\rangle$ basis becomes

$$H = \begin{pmatrix} \overline{E} + q\mathcal{E}s\cos\omega t & -A \\ -A & \overline{E} - q\mathcal{E}s\cos\omega t \end{pmatrix}. \tag{5.59}$$

Transform this Hamiltonian from the $|\pm\rangle$ basis to the basis provided by the states of well-defined parity $|\mathrm{e}\rangle$ and $|\mathrm{o}\rangle$ (where $|\mathrm{e}\rangle = (|+\rangle + |-\rangle)/\sqrt{2}$, etc). Writing

$$|\psi\rangle = a_\mathrm{e}(t)\mathrm{e}^{-\mathrm{i}E_\mathrm{e}t/\hbar}|\mathrm{e}\rangle + a_\mathrm{o}(t)\mathrm{e}^{-\mathrm{i}E_\mathrm{o}t/\hbar}|\mathrm{o}\rangle, \tag{5.60}$$

show that the equations of motion of the expansion coefficients are

$$\begin{aligned} \frac{\mathrm{d}a_\mathrm{e}}{\mathrm{d}t} &= -\mathrm{i}\Omega a_\mathrm{o}(t)\Big(\mathrm{e}^{\mathrm{i}(\omega-\omega_0)t} + \mathrm{e}^{-\mathrm{i}(\omega+\omega_0)t}\Big) \\ \frac{\mathrm{d}a_\mathrm{o}}{\mathrm{d}t} &= -\mathrm{i}\Omega a_\mathrm{e}(t)\Big(\mathrm{e}^{\mathrm{i}(\omega+\omega_0)t} + \mathrm{e}^{-\mathrm{i}(\omega-\omega_0)t}\Big), \end{aligned} \tag{5.61}$$

where $\Omega \equiv q\mathcal{E}s/2\hbar$ and $\omega_0 = (E_\mathrm{o} - E_\mathrm{e})/\hbar$. Explain why in the case of a maser the exponentials involving $\omega + \omega_0$ a can be neglected so the equations of motion become

$$\frac{\mathrm{d}a_\mathrm{e}}{\mathrm{d}t} = -\mathrm{i}\Omega a_\mathrm{o}(t)\mathrm{e}^{\mathrm{i}(\omega-\omega_0)t} \quad ; \quad \frac{\mathrm{d}a_\mathrm{o}}{\mathrm{d}t} = -\mathrm{i}\Omega a_\mathrm{e}(t)\mathrm{e}^{-\mathrm{i}(\omega-\omega_0)t}. \tag{5.62}$$

Solve the equations by multiplying the first equation by $\mathrm{e}^{-\mathrm{i}(\omega-\omega_0)t}$ and differentiating the result. Explain how the solution describes the decay of a population of molecules that are initially all in the higher energy level. Compare your solution to the result of setting $\omega = \omega_0$ in (5.62).

5.11 Show that the amplitude C of the outgoing wave in equations (5.32) is

$$C = \left(\frac{2Q}{Q+1}\right)\frac{A_+\mathrm{e}^{-2\mathrm{i}ka}}{\cos 2Ka - \mathrm{i}(k/K)\sin 2Ka}. \tag{5.63}$$

Explain physically the amplitude and phase of this expression when $2Ka = r\pi$ with $r = 1, 2, \ldots$.

5.12* Particles of mass m and momentum $\hbar k$ at $x < -a$ move in the potential

$$V(x) = V_0 \begin{cases} 0 & \text{for } x < -a \\ \frac{1}{2}[1 + \sin(\pi x/2a)] & \text{for } |x| < a \\ 1 & \text{for } x > a, \end{cases} \tag{5.64}$$

where $V_0 < \hbar^2k^2/2m$. Numerically reproduce the reflection probabilities plotted Figure 5.15 as follows. Let $\psi_i \equiv \psi(x_j)$ be the value of the wavefunction at $x_j = j\Delta$, where Δ is a small increment in the x coordinate. From the TISE show that

$$\psi_j \simeq (2 - \Delta^2k^2)\psi_{j+1} - \psi_{j+2}, \tag{5.65}$$

where $k \equiv \sqrt{2m(E-V)}/\hbar$. Determine ψ_j at the two grid points with the largest values of x from a suitable boundary condition, and use the recurrence relation (5.65) to determine ψ_j at all other grid points. By matching the values of ψ at the points with the smallest values of x to a sum of sinusoidal waves, determine the probabilities required for the figure. Be sure to check the accuracy of your code when $V_0 = 0$, and in the general case explicitly check that your results are consistent with equal fluxes of particles towards and away from the origin.

Equation (11.40) gives an analytical approximation for ψ in the case that there is negligible reflection. Compute this approximate form of ψ and compare it with your numerical results for larger values of a.

5.13* In this problem we obtain an analytic estimate of the energy difference between the even- and odd-parity states of a double square well. Show that for large θ, $\coth\theta - \tanh\theta \simeq 4\mathrm{e}^{-2\theta}$. Next letting δk be the difference between the k values that solve

$$\tan\left[r\pi - k(b-a)\right]\sqrt{\frac{W^2}{(ka)^2}-1} = \begin{cases} \coth\left(\sqrt{W^2-(ka)^2}\right) & \text{even parity} \\ \tanh\left(\sqrt{W^2-(ka)^2}\right) & \text{odd parity,} \end{cases} \tag{5.66a}$$

where

$$W \equiv \sqrt{\frac{2mV_0a^2}{\hbar^2}} \tag{5.66b}$$

for given r in the odd- and even-parity cases, deduce that

$$\left\{\left[\left(\frac{W^2}{(ka)^2}-1\right)^{1/2} + \left(\frac{W^2}{(ka)^2}-1\right)^{-1/2}\right](b-a) + \frac{1}{k}\left(1-\frac{(ka)^2}{W^2}\right)^{-1}\right\}\delta k \\ \simeq -4\exp\left[-2\sqrt{W^2-(ka)^2}\right]. \tag{5.67}$$

Hence show that when $W \gg 1$ the fractional difference between the energies of the ground and first excited states is

$$\frac{\delta E}{E} \simeq \frac{-8a}{W(b-a)}\,\mathrm{e}^{-2W\sqrt{1-E/V_0}}. \tag{5.68}$$

6

Composite systems

Systems often consist of more than one part. For example a hydrogen atom consists of a proton and an electron, and a diamond consists of a very large number of carbon atoms. In these examples of **composite systems** there is significant physical interaction between the component parts of the system – the electron moves in the electromagnetic field of the proton, and electromagnetic forces act between the atoms in a diamond. But in principle there need be no physical interaction between the parts of a composite system: it is enough that we consider the sum of the parts to constitute a single system. For example 'quantum cryptography' exploits correlations between widely separated photons that are not interacting with each other, and in §7.5 we shall study a system that consists of two completely unconnected gyros that happen to be in the same box. Even in classical physics specifying the state of such a system is a complex business because in general there will be **correlations** between the parts of the system: the probability for obtaining a certain value for an observable of one subsystem depends on the state of the other subsystem. In quantum mechanics correlations arise through quantum interference between various states of the system, with the result that correlations are sometimes associated with unexpected and sometimes puzzling phenomena.

In §6.1 we extend the formalism of quantum mechanics to composite systems. We introduce the concept of 'quantum entanglement', which is how correlations between the different parts of a composite system are represented in quantum mechanics, and we find that subsystems have a propensity to become entangled. In §6.1.4 we discuss a thought experiment with entangled particles that Einstein believed demonstrated that quantum mechanics is merely an incomplete substitute for a deeper theory. Subsequently an experiment of this type was carried out and the results showed that a theory of the type sought by Einstein must conflict with experiments. In §6.2 we introduce the principal ideas of quantum computing, which is the focus of much current experimental work and has the potential to revolutionise computational mathematics with major implications for the many aspects of our civilisation that rely on cryptography. In §6.3 we introduce the operator that enables us to drop unrealistic assumptions about our level of knowledge of the states of quantum systems and introduce the key concept of entropy. In §6.4 we show that thermodynamics arises naturally from quantum mechanics. In §6.5 we come clean about the intellectual black hole that lurks at the heart of quantum mechanics: the still unresolved problem of measurement.

At several points in the chapter we encounter fundamental questions about quantum mechanics with which experimental and theoretical physicists are currently wrestling. It is a remarkable feature of quantum mechanics that already the sixth chapter of an introduction to the subject can bring students to the frontier of human understanding.

6.1 Composite systems

Once we understand how to combine two systems A and B to make a composite system AB, we will be in a position to build up systems of arbitrary complexity, because we will be able to combine the system AB with some other system C to make a system ABC, and so on indefinitely. So we now consider what is involved in forming AB out of A and B.

Suppose $\{|\mathrm{A};i\rangle\}$ and $\{|\mathrm{B};j\rangle\}$ are sets of states of A and B, respectively. Then the symbolic product $|\mathrm{A};i\rangle|\mathrm{B};j\rangle$ is used to denote that state of the composite system in which A is in the state $|\mathrm{A};i\rangle$ and B is in the state $|\mathrm{B};j\rangle$: clearly this *is* a well defined state of AB. We express this fact by writing

$$|\mathrm{AB};i,j\rangle = |\mathrm{A};i\rangle|\mathrm{B};j\rangle, \tag{6.1a}$$

where the label of the ket before the semicolon indicates what system is having its state specified, and the label after the semicolon enumerates the states. The Hermitian adjoint of equation (6.1a) is

$$\langle\mathrm{AB};i,j| = \langle\mathrm{A};i|\langle\mathrm{B};j|, \tag{6.1b}$$

and we define the product of a bra and a ket of AB by the rule

$$\langle\mathrm{AB};i',j'|\mathrm{AB};i,j\rangle = \langle\mathrm{A};i'|\mathrm{A};i\rangle\langle\mathrm{B};j'|\mathrm{B};j\rangle. \tag{6.2}$$

This rule is well defined because the right side is simply a product of two complex numbers. It is a physically sensible rule because it implies that the probability that AB is in the state $i'j'$ is the product of the probability that A is in state i' and B is in state j':

$$\begin{aligned} p(\mathrm{AB};i',j') &= |\langle\mathrm{AB};i',j'|\mathrm{AB};i,j\rangle|^2 = |\langle\mathrm{A};i'|\mathrm{A};i\rangle|^2|\langle\mathrm{B};j'|\mathrm{B};j\rangle|^2 \\ &= p(\mathrm{A};i')p(\mathrm{B};j'). \end{aligned} \tag{6.3}$$

Any state of AB that like (6.2) can be written as a product of a state of A and a state of B is rather special. To see this, we consider the simplest non-trivial example, in which both A and B are two-state systems. Let $|+\rangle$ and $|-\rangle$ be the two basis states $|\mathrm{A};i\rangle$ of A and let $|\uparrow\rangle$ and $|\downarrow\rangle$ be the two basis states $|\mathrm{B};j\rangle$ of B – we shall call these the 'up' and 'down' states of B. We use these basis states to expand the states of the subsystems:

$$|\mathrm{A}\rangle = a_-|-\rangle + a_+|+\rangle \quad ; \quad |\mathrm{B}\rangle = b_\downarrow|\downarrow\rangle + b_\uparrow|\uparrow\rangle, \tag{6.4}$$

so the state $|\mathrm{AB}\rangle = |\mathrm{A}\rangle|\mathrm{B}\rangle$ of AB can be written

$$\begin{aligned} |\mathrm{AB}\rangle &= (a_-|-\rangle + a_+|+\rangle)\,(b_\downarrow|\downarrow\rangle + b_\uparrow|\uparrow\rangle,) \\ &= a_-b_\downarrow|-\rangle|\downarrow\rangle + a_-b_\uparrow|-\rangle|\uparrow\rangle + a_+b_\downarrow|+\rangle|\downarrow\rangle + a_+b_\uparrow|+\rangle|\uparrow\rangle. \end{aligned} \tag{6.5}$$

The coefficients in this expansion are the amplitudes for particular events – for example $a_-b_\downarrow$ is the amplitude that A will be found to be minus *and* B will be found to be down. From them we obtain a relation between the probabilities of finding A to be in its plus state and B to be either up or down:

$$\frac{p_{+\uparrow}}{p_{+\downarrow}} = \frac{|b_\uparrow|^2}{|b_\downarrow|^2}. \tag{6.6}$$

Now by Bayes' theorem, the probability of finding B to be up given that A is plus is

$$p(\mathrm{B};\uparrow|\mathrm{A};+) = \frac{p_{+\uparrow}}{p(\mathrm{A};+)} = \frac{p_{+\uparrow}}{p_{+\uparrow}+p_{+\downarrow}} = \frac{1}{1+p_{+\downarrow}/p_{+\uparrow}}. \tag{6.7}$$

With equation (6.6) this simplifies to

$$p(\mathrm{B};\uparrow|\mathrm{A};+) = \frac{1}{1+|b_\downarrow/b_\uparrow|^2}. \tag{6.8}$$

The key thing is that the right side of this expression makes no reference to subsystem A. Evidently, when the state $|\mathrm{AB}\rangle$ of the composite system can be written as a product $|\mathrm{A}\rangle|\mathrm{B}\rangle$ of states of the subsystems, the probability of finding B to be up is independent of the state of A. That is, the two subsystems are **uncorrelated** or **statistically independent**. Usually the states of subsystems *are* correlated and then the state of AB cannot be expressed as a simple product $|\mathrm{A}\rangle|\mathrm{B}\rangle$.

For example, suppose we have two vertical gear wheels, A with N_A teeth and B with N_B teeth. Then the state of A is specified by giving the amplitudes a_i that the i^{th} tooth is on top of the wheel. The state of B is similarly specified by the amplitudes b_j for each of its teeth to be uppermost. However, if both wheels are members of the same train of gears (as in a clock), the probability that the j^{th} tooth of B is on top will depend on which tooth of A is uppermost. When the orientations of the wheels are correlated in this way, each of the $N_\mathrm{A}N_\mathrm{B}$ configurations of the pair of wheels has an independent probability, p_{ij}. Specifically, when $N_\mathrm{A} = N_\mathrm{B}$, p_{ij} will vanish except when $i = j$. If these gear wheels are uncorrelated because they are not meshed together, we need to specify only the $N_\mathrm{A}+N_\mathrm{B}$ amplitudes a_i and b_j. Once the wheels become correlated as a result of their teeth meshing, we have to specify $N_\mathrm{A}N_\mathrm{B}$ amplitudes, one for each probability p_{ij}.

We now assume that the sets $\{|\mathrm{A};i\rangle\}$ and $\{|\mathrm{B};j\rangle\}$ are complete for their respective systems and show that the set of states given by equation (6.1a) for all possible values of i, j is then a complete set of states for the composite system. That is, *any* state $|\mathrm{AB};\psi\rangle$ of AB can be written

$$|\mathrm{AB};\psi\rangle = \sum_{ij} c_{ij}|\mathrm{AB};i,j\rangle = \sum_{ij} c_{ij}|\mathrm{A};i\rangle|\mathrm{B};j\rangle. \tag{6.9}$$

The proof involves supposing on the contrary that there is a state $|\mathrm{AB};\phi\rangle$ of AB that cannot be expressed in the form (6.9). We construct the object

$$|\mathrm{AB};\chi\rangle \equiv |\mathrm{AB};\phi\rangle - \sum_{ij} c_{ij}|\mathrm{AB};i,j\rangle \quad \text{where} \quad c_{ij} = \langle \mathrm{AB};i,j|\mathrm{AB};\phi\rangle. \tag{6.10}$$

This object cannot vanish or $|\mathrm{AB};\phi\rangle$ *would* be of the form (6.9). But when AB is in this state, the amplitude for subsystem A to be in *any* of the states $|\mathrm{A};i\rangle$ vanishes:

$$\sum_j \left(\langle \mathrm{A};i|\langle \mathrm{B};j|\right)|\mathrm{AB};\chi\rangle = 0. \tag{6.11}$$

This conclusion is absurd because the set $\{|\mathrm{A};i\rangle\}$ is by hypothesis complete, so the hypothesised state $|\mathrm{AB};\phi\rangle$ cannot exist. Thus we have shown that a general state of AB is specified by $N_\mathrm{A}N_\mathrm{B}$ amplitudes, just as the argument about gear wheels suggested.

This result implies that the number of amplitudes required to specify the state of a composite system grows exceedingly rapidly with the complexity of the subsystems – for example, if $N_\mathrm{A} = N_\mathrm{B} = 1000$, a million amplitudes are required to specify a general state of AB. By contrast only 2000 amplitudes are required to specify a product state $|\mathrm{AB}\rangle = |\mathrm{A}\rangle|\mathrm{B}\rangle$ because the form of such a state automatically sets to zero all correlations between the subsystems. For a general state a large number of amplitudes are required to specify these correlations.

Even when a state of AB is given by an expansion of the form (6.9) that involves $N_\mathrm{A}N_\mathrm{B}$ amplitudes, the states of A and B may not be correlated. To see this let $|\mathrm{A};\psi\rangle = \sum_{i=1}^{N_\mathrm{A}} a_i|\mathrm{A};i\rangle$ be the state of the subsystem A and let $|\mathrm{B};\phi\rangle = \sum_{j=1}^{N_\mathrm{B}} b_j|\mathrm{B};j\rangle$ be the state of subsystem B. Then the state of the composite system AB is

$$|\mathrm{AB};\chi\rangle \equiv |\mathrm{A};\psi\rangle|\mathrm{B};\phi\rangle = \sum_{ij} a_i b_j |\mathrm{A};i\rangle|\mathrm{B};j\rangle. \tag{6.12}$$

The right side of this equation is identical to the right side of equation (6.9) except that c_{ij} has been replaced with $a_i b_j$. Thus equation (6.12) is an instance of the general expansion (6.9), but it is a very *special* instance: in general the expansion coefficients c_{ij}, which can be thought of as the entries in an $N_\mathrm{A} \times N_\mathrm{B}$ matrix, cannot be written as the product of an N_A-dimensional vector with entries a_i and an N_B-dimensional vector b_j. To see that this is so, consider the ratio $c_{ij}/c_{ij'}$ of the matrix elements in the same row but different columns. When c_{ij} can be expressed as the product of two vectors, we have

$$\frac{c_{ij}}{c_{ij'}} = \frac{a_i b_j}{a_i b_{j'}} = \frac{b_j}{b_{j'}}, \tag{6.13}$$

so this ratio is independent of i. That is, when the state of AB can be written as the product of a state of A and a state of B, the expansion coefficients c_{ij} are restricted such that every row of the matrix that they form is a multiple of the top row. Similarly, in this case every column is a multiple of the leftmost column (Problem 6.2).

When the state of AB cannot be written as the product of a state of A and a state of B, we say that the subsystems A and B are **entangled**. As we have seen, the observables of entangled systems are correlated, so we could as well say that the subsystems are correlated. It is remarkable that correlations between subsystems, which are as evident in classical physics as in quantum mechanics, arise in quantum mechanics through the quintessentially quantum phenomenon of the addition of quantum *amplitudes*: states of AB in which subsystems A and B are correlated are expressed as linear combinations of states in which A and B are uncorrelated. The use of the word ‘entanglement’ reminds us that correlations arise through an intertwining of states that is inherently quantum-mechanical and without classical analogue.

It may help to clarify these ideas if we apply them to a hydrogen atom. We work in the position representation, so we require the amplitude

$$\psi(\mathbf{x}_\mathrm{e}, \mathbf{x}_\mathrm{p}) = \langle \mathbf{x}_\mathrm{e}, \mathbf{x}_\mathrm{p}|\psi\rangle \tag{6.14}$$

to find the electron near $\mathbf{x}_\mathrm{e}$ and the proton near $\mathbf{x}_\mathrm{p}$. Suppose that we have states

$$u_i(\mathbf{x}_\mathrm{e}) = \langle \mathbf{x}_\mathrm{e}|u_i\rangle \quad \text{and} \quad U_j(\mathbf{x}_\mathrm{p}) = \langle \mathbf{x}_\mathrm{p}|U_j\rangle \tag{6.15}$$

that form complete sets for the electron and the proton, respectively. Then for any state of the atom, $|\psi\rangle$, there are numbers c_{ij} such that

$$|\psi\rangle = \sum_{ij} c_{ij}|u_i\rangle|U_j\rangle. \tag{6.16}$$

Box 6.1: Classical correlations

It's instructive to consider how we can represent correlations between two classical systems A and B. Let's assume that each system has a finite number N of discrete states – they might be digital counters on an instrument panel. Then there are N^2 probabilities c_{jk} to specify.

We can specify the state of A by giving N probabilities a_j and similarly the state of B can be specified by probabilities b_k. We might choose to express these in terms of their discrete Fourier transforms $\hat{a}_\alpha$ and $\hat{b}_\beta$, so

$$a_j = \sum_{\alpha=0}^{N-1} \hat{a}_\alpha \mathrm{e}^{2\pi\mathrm{i}\alpha j/N} \quad ; \quad b_k = \sum_{\beta=0}^{N-1} \hat{b}_\beta \mathrm{e}^{2\pi\mathrm{i}\beta k/N}$$

If A and B were uncorrelated, so $c_{jk} = a_j b_k$, the state of AB could be written

$$c_{jk} = \sum_{\alpha\beta} \hat{c}_{\alpha\beta} \mathrm{e}^{2\pi\mathrm{i}(\alpha j+\beta k)/N}, \tag{1}$$

where

$$\hat{c}_{\alpha\beta} = \hat{a}_\alpha b_\beta. \tag{2}$$

In the presence of correlations we can still represent c_{jk} as the double Fourier sum (1) but then $\hat{c}_{\alpha\beta}$ will not be given by the product of equation (2). Thus the mathematical manifestation of classical correlations can be very similar to quantum entanglement. The big difference is that in the classical case the expansion coefficients have no physical interpretation: the basis functions used for expansion (here the circular functions $\mathrm{e}^{2\pi\mathrm{i}\alpha j/N}$) and the expansion coefficients $\hat{a}_\alpha$ etc., will not be non-negative so they cannot be interpreted as probability distributions. In quantum mechanics these quantities acquire physical interpretations. Moreover, the final probabilities, being obtained by mod-squaring a sum like that of equation (1), involve quantum interference between different terms in the sum.

Multiplying through by $\langle \mathbf{x}_\mathrm{e}, \mathbf{x}_\mathrm{p}|$ we obtain

$$\psi(\mathbf{x}_\mathrm{e}, \mathbf{x}_\mathrm{p}) = \sum_{ij} c_{ij} u_i(\mathbf{x}_\mathrm{e}) U_j(\mathbf{x}_\mathrm{p}). \tag{6.17}$$

The product of u_i and U_j on the right is no longer symbolic: it is an ordinary product of complex numbers. The quantity c_{ij} is the amplitude to find the electron in the state $|u_i\rangle$ and the proton in the state $|U_j\rangle$.

6.1.1 Collapse of the wavefunction

Consider again the composite system we introduced above in which both A and B are two-state systems, with $|-\rangle$ and $|+\rangle$ constituting a basis for A and $|\downarrow\rangle$ and $|\uparrow\rangle$ constituting a basis for B. Let AB be in the entangled state

$$|\mathrm{AB}\rangle = a|+\rangle|\uparrow\rangle + |-\rangle(b|\uparrow\rangle + c|\downarrow\rangle), \tag{6.18a}$$

where b and c are given complex numbers. Then if a measurement of subsystem A is made and it yields +, the state of AB after the system is

$$|\mathrm{AB}\rangle = |+\rangle|\uparrow\rangle. \tag{6.18b}$$

Conversely, if the measurement of A yields −, the state of AB after the measurement is

$$|\mathrm{AB}\rangle = \frac{1}{\sqrt{|b|^2 + |c|^2}}\,|-\rangle(b|\uparrow\rangle + c|\downarrow\rangle). \tag{6.18c}$$

These rules are extensions of the usual collapse hypothesis, which we introduced in §1.4: there we had a single system and we stated that when a measurement is made, the state of the system collapses from a linear combination of states that are each possible outcomes of the measurement to the particular state that corresponds to the value of the observable actually measured. That is

$$|\psi\rangle = \sum_i a_i|i\rangle \quad \rightarrow \quad |\psi\rangle = |3\rangle, \quad \text{say.} \tag{6.19}$$

The new twist in equations (6.18) is that when we expand the state of a composite system as a linear combination of states of the subsystem we propose to measure, the coefficients of those states are states of the other subsystem rather than amplitudes, and these states are the ones the second system will be in after the first system has been measured. Consequently, the amplitudes we obtain for a subsequent measurement of the second subsystem depend on the outcome of the first measurement: if measurement of A yields $+$, then from (6.18b), a measurement of B is certain to yield $\uparrow$, while if the measurement of A yields $-$, subsequent measurement of B will yield $\downarrow$ with probability $1/(|b/c|^2+1)$.

6.1.2 Operators for composite systems

While the law of multiplication of probabilities leads to the kets of subsystems being multiplied, we *add* the operators of subsystems. For example, if A and B are both free particles, then the Hamiltonian operator of the composite system is

$$H_{\rm AB} = H_{\rm A} + H_{\rm B} = \frac{p_{\rm A}^2}{2m_{\rm A}} + \frac{p_{\rm B}^2}{2m_{\rm B}}. \tag{6.20}$$

In this simple example there is no physical interaction between the parts of the system, with the consequence that the Hamiltonian splits into a part that depends only on the operators of A, and a part that depends only on operators of B. When there is a physical connection between the systems, there will be an additional part of the Hamiltonian, the **interaction Hamiltonian** that depends on operators belonging to both systems. For example, if both particles bear electrostatic charge Q, the interaction Hamiltonian

$$H_{\rm int} = \frac{Q^2}{4\pi\epsilon_0|\mathbf{x}_{\rm A} - \mathbf{x}_{\rm B}|} \tag{6.21}$$

should be added to $H_{\rm A} + H_{\rm B}$ to form $H_{\rm AB}$. For the rest of this subsection we assume for simplicity that there is no dynamical interaction between the subsystems.

When an operator acts on a ket that is a product of one describing A and one describing B, kets that belong to the other system stand idly by as if they were mere complex numbers. For example

$$p_{\rm B}|{\rm A};i\rangle|{\rm B};j\rangle = |{\rm A};i\rangle\big(p_{\rm B}|{\rm B};j\rangle\big) \tag{6.22}$$

so

$$\begin{aligned}\langle{\rm A};i'|\langle{\rm B};j'|(H_{\rm A}+H_{\rm B})|{\rm A};i\rangle|{\rm B};j\rangle &= \langle{\rm A};i'|H_{\rm A}|{\rm A};i\rangle\langle{\rm B};j'|{\rm B};j\rangle \\ &\quad + \langle{\rm A};i'|{\rm A};i\rangle\langle{\rm B};j'|H_{\rm B}|{\rm B};j\rangle \\ &= \langle{\rm A};i'|H_{\rm A}|{\rm A};i\rangle\delta_{jj'} + \delta_{ii'}\langle{\rm B};j'|H_{\rm B}|{\rm B};j\rangle.\end{aligned} \tag{6.23}$$

When we set $i' = i$ and $j' = j$ we obtain the expectation value of $H_{\rm AB}$ when the system is in the state $|{\rm A};i\rangle|{\rm B};j\rangle$. This is easily seen to be just the sum of the expectation values of the energies of the two free particles, as one would expect.

We shall several times have to find the eigenvalues and eigenkets of an operator such as $H_{\rm AB}$ that is the sum of operators $H_{\rm A}$ and $H_{\rm B}$ that belong to completely different subsystems. Every operator of subsystem A commutes with *every* operator of subsystem B. Consequently when $H_{\rm AB}$ is given by equation (6.20),

$$[H_{\rm AB}, H_{\rm A}] = [H_{\rm A} + H_{\rm B}, H_{\rm A}] = 0. \tag{6.24}$$

That is, when there is no physical interaction between the subsystems, so $H_{\rm AB}$ is just the sum of the Hamiltonians of the individual systems, $H_{\rm AB}$ commutes with both individual Hamiltonians. It follows that in this case there is a complete set of mutual eigenkets of $H_{\rm AB}$, $H_{\rm A}$ and $H_{\rm B}$. Let $\{|{\rm A}; i\rangle\}$ be a complete set of eigenkets of $H_{\rm A}$ with eigenvalues $E_i^{\rm A}$, and let $\{|{\rm B}; i\rangle\}$ be a complete set of eigenkets of $H_{\rm B}$ with eigenvalues $E_j^{\rm B}$. Then it is trivial to check that the states $|{\rm AB}; i, j\rangle \equiv |{\rm A}; i\rangle|{\rm B}; j\rangle$ are eigenkets of $H_{\rm AB}$ with eigenvalues $E_i^{\rm A} + E_j^{\rm B}$. Moreover, we showed above that these product kets form a complete set. So the states $|{\rm AB}; ij\rangle$ form a complete set of mutual eigenkets of $H_{\rm AB}$, $H_{\rm A}$ and $H_{\rm B}$. In the position representation this result becomes the statement that the wavefunctions

$$\psi_{ij}^{\rm AB}(\mathbf{x}_{\rm A}, \mathbf{x}_{\rm B}) \equiv \langle \mathbf{x}_{\rm A}, \mathbf{x}_{\rm B}|{\rm AB}; ij\rangle = u_i^{\rm A}(\mathbf{x}_{\rm A}) u_j^{\rm B}(\mathbf{x}_{\rm B}) \tag{6.25}$$

form a complete set of mutual eigenfunctions for the three operators. That is, if we have a composite system with a Hamiltonian that is simply the sum of the Hamiltonians of the parts, we can assume that the eigenfunctions of the whole system's Hamiltonian are simply products of eigenfunctions of the individual component Hamiltonians.

It is instructive to write the TDSE for a composite system formed by two non-interacting subsystems:

$$\begin{aligned} {\rm i}\hbar\frac{\partial|{\rm AB}\rangle}{\partial t} &= {\rm i}\hbar\frac{\partial}{\partial t}(|{\rm A}\rangle|{\rm B}\rangle) = {\rm i}\hbar\left(\frac{\partial|{\rm A}\rangle}{\partial t}|{\rm B}\rangle + |{\rm A}\rangle\frac{\partial|{\rm B}\rangle}{\partial t}\right) \\ &= (H_{\rm A}|{\rm A}\rangle)|{\rm B}\rangle + |{\rm A}\rangle(H_{\rm B}|{\rm B}\rangle) = (H_{\rm A} + H_{\rm B})|{\rm A}\rangle|{\rm B}\rangle \\ &= H_{\rm AB}|{\rm AB}\rangle. \end{aligned} \tag{6.26}$$

Thus we have been able to derive the TDSE for the composite system from the TDSE for each subsystem. Notice that the physically evident rule for adding the Hamiltonians of the subsystem emerges as a consequence of the ket for the whole system being a product of the kets of the subsystems and the usual rule for differentiating a product.

6.1.3 Development of entanglement

Entangled is an appropriate name because subsystems are as prone to become entangled as is the line of a kite. To justify this statement, we consider the dynamical evolution of a composite system AB. Without loss of generality we can use basis states that satisfy the TDSEs of the isolated subsystems. That is, the we may assume that the states $|{\rm A}; i\rangle$, etc, satisfy

$${\rm i}\hbar\frac{\partial|{\rm A}; i\rangle}{\partial t} = H_{\rm A}|{\rm A}; i\rangle \quad \text{and} \quad {\rm i}\hbar\frac{\partial|{\rm B}; j\rangle}{\partial t} = H_{\rm B}|{\rm B}; j\rangle. \tag{6.27}$$

A general state of the composite system is

$$|{\rm AB}\rangle = \sum_{ij} c_{ij}|{\rm A}; i\rangle|{\rm B}; j\rangle, \tag{6.28}$$

where the expansion coefficients c_{ij} are all functions of time. The Hamiltonian of the composite system can be written

$$H_{\rm AB} = H_{\rm A} + H_{\rm B} + H_{\rm int}, \tag{6.29}$$

where the interaction Hamiltonian $H_{\rm int}$ is the part of $H_{\rm AB}$ that contains operators belonging to both subsystems (cf eq. 6.21). Substituting this expression for $H_{\rm AB}$ and the expansion (6.28) into the TDSE for the composite system (eq. 6.26), we find

$$\begin{aligned}\mathrm{i}\hbar\frac{\partial|\mathrm{AB}\rangle}{\partial t}&=\mathrm{i}\hbar\sum_{ij}\left\{\frac{\mathrm{d}c_{ij}}{\mathrm{d}t}|\mathrm{A};i\rangle|\mathrm{B};j\rangle+c_{ij}\left(\frac{\partial|\mathrm{A};i\rangle}{\partial t}|\mathrm{B};j\rangle+|\mathrm{A};i\rangle\frac{\partial|\mathrm{B};j\rangle}{\partial t}\right)\right\}\\&=\sum_{ij}c_{ij}\left\{(H_{\rm A}|\mathrm{A};i\rangle)|\mathrm{B};j\rangle+|\mathrm{A};i\rangle(H_{\rm B}|\mathrm{B};j\rangle)+H_{\rm int}|\mathrm{A};i\rangle|\mathrm{B};j\rangle\right\}.\end{aligned}\tag{6.30}$$

After using equations (6.27) to cancel terms, this simplifies to

$$\mathrm{i}\hbar\sum_{ij}\frac{\mathrm{d}c_{ij}}{\mathrm{d}t}|\mathrm{A};i\rangle|\mathrm{B};j\rangle=\sum_{ij}c_{ij}H_{\rm int}|\mathrm{A};i\rangle|\mathrm{B};j\rangle,\tag{6.31}$$

which states that the time evolution of the expansion coefficients is entirely driven by the interaction Hamiltonian. In particular, if there is no coupling between the systems ($H_{\rm int}=0$), the c_{ij} are constant, so if the systems are initially unentangled, they remain so.

By multiplying equation (6.31) through by $\langle\mathrm{A};k|\langle\mathrm{B};l|$ we obtain an equation that is most conveniently written

$$\mathrm{i}\hbar\frac{\mathrm{d}c_{kl}}{\mathrm{d}t}=\sum_{ij}c_{ij}\langle\mathrm{AB};kl|H_{\rm int}|\mathrm{AB};ij\rangle.\tag{6.32}$$

Let's suppose that all the matrix elements in this equation vanish except an element $\langle\mathrm{AB};k_0l_0|H_{\rm int}|\mathrm{AB};k_0l_0\rangle$ which lies on the diagonal. Then only $c_{k_0l_0}$ will have non-vanishing time derivative, so the condition for the subsystems to be unentangled, namely that $c_{ij}/c_{ij'}$ is independent of i, which is initially satisfied, will soon be violated by the ratio $c_{k_0l_0}/c_{k_0j}$ for $j\neq l_0$. Careful consideration of what happens when there are several non-vanishing matrix elements leads to the same conclusion: almost any coupling between subsystems will cause them to become entangled from an unentangled initial condition.

This result is not surprising physically: a coupling makes the motion of one system dependent on the state of the other. So after some time the state that the second system has reached depends on the state of the first system, which is just to say that the two systems have become correlated or entangled.

6.1.4 Einstein–Podolski–Rosen experiment

In 1935 A. Einstein, B. Podolski and N. Rosen (**EPR** for short) proposed[1] an experiment with entangled particles that they argued would demonstrate that quantum mechanics is an incomplete theory in the sense that to specify the state of a physical system you need to know the values taken by **hidden variables** that quantum mechanics does not consider. In 1964 J.S. Bell showed[2] that for a similar experiment quantum mechanics makes predictions that are incompatible with the existence of hidden variables. In 1972 an experiment of this type was successfully carried out[3] and its results were found to vindicate quantum mechanics. We now describe Bell's formulation of the experiment and discuss its implications.

A nucleus decays from a state that has no spin to another spinless state by emitting an electron and a positron. The nucleus is at rest both before and

---

[1] E. Einstein, B. Podolski & N. Rosen, Phys. Rev. **47**, 777 (1935)

[2] J.S. Bell, Phyics, 1, 195 (1964)

[3] S.J. Freedman & J.F. Clauser, Phys. Rev. L., 28, 938 (1972)

after the decay, so the electron and positron move away in opposite directions with equal speeds. As we saw in §1.3.5, electrons and positrons are spinning particles so they each carry some spin angular momentum away from the nucleus. Since the nucleus is at all times without angular momentum, the angular momenta of the electron and positron must be equal and opposite. At some distance from the decaying nucleus Alice detects the electron and measures the component of its spin in the direction of her choice, **a**. As we saw in §1.3.5, the result of this measurement will be either $+\frac{1}{2}$ or $-\frac{1}{2}$. Meanwhile Bob, who sits a similar distance from the nucleus to Alice, detects the positron and measures its spin in the direction of his choice, **b**.

After Alice has obtained $+\frac{1}{2}$ on measuring the spin along **a** she thinks: "If Bob measures along **a** too, he *must* measure $-\frac{1}{2}$. But if Bob measures along some other vector **b**, I cannot be certain what value he will get, but he isn't likely to get $+\frac{1}{2}$ if **b** is only slightly inclined to my vector **a**, that is, if $1-\mathbf{a}\cdot\mathbf{b} \ll 1$." Alice can see that conservation of angular momentum implies that the results obtained by Bob and herself must be correlated. Let's put this argument on a quantitative basis.

In §7.5.1 we shall see that because the system formed by the electron-positron pair has no net angular momentum, its state can be written

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|\mathrm{e}+\rangle|\mathrm{p}-\rangle - |\mathrm{e}-\rangle|\mathrm{p}+\rangle\right). \tag{6.33}$$

Here $|\mathrm{e}+\rangle$ is the state in which the component of the electron's spin along the z-axis is certain to be $+\frac{1}{2}$, and similarly for $|\mathrm{p}\pm\rangle$, etc. We are free to orient the z-axis parallel to Alice's choice of direction **a**, so we do this. When Alice obtains $+\frac{1}{2}$, she collapses the system's state into

$$|\psi'\rangle = |\mathrm{e}+\rangle|\mathrm{p}-\rangle. \tag{6.34}$$

Before Alice's measurement, when the state was given by equation (6.33), the amplitude for a measurement of the positron's spin along **a** to yield $+\frac{1}{2}$ was $1/\sqrt{2}$, but after the measurement equation (6.34) shows that it vanishes, just as Alice reasoned it would. To find the amplitude for Bob to measure for the positron $+\frac{1}{2}$ along another vector **b**, we recall equation (1.34a) from §1.3.5:

$$|+,\mathbf{b}\rangle = \sin(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi/2}|\mathrm{p}-\rangle + \cos(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi/2}|\mathrm{p}+\rangle, \tag{6.35}$$

where θ and ϕ are the polar angles that give the orientation of **b** in a system in which **a** is along the z-axis. In particular

$$\cos\theta = \mathbf{a}\cdot\mathbf{b}. \tag{6.36}$$

Given that after Alice's measurement the positron is certainly in the state $|\mathrm{p}-\rangle$, it follows from equation (6.35) that the amplitude for Bob to measure $+\frac{1}{2}$ along his chosen direction is $\langle +,\mathbf{b}|\mathrm{p}-\rangle = \sin(\theta/2)\mathrm{e}^{-\mathrm{i}\phi/2}$. Mod-squaring this amplitude we find that the probability that Bob measures $+\frac{1}{2}$ is

$$P_{\mathrm{B}}(+|\mathrm{A}+) = \sin^2(\theta/2), \tag{6.37}$$

which is small when $\mathbf{a} \simeq \mathbf{b}$ as Alice predicted. So quantum mechanics is consistent with common sense.

We have supposed that Alice measures first, but if the electron and positron are moving relativistically, a light signal sent to Bob by Alice when she made her measurement would not have arrived at Bob when he made *his* measurement, and vice versa. In these circumstances the theory of relativity teaches us that the order in which the measurements are made depends on the velocity of the observer who is judging the matter. Consequently, for consistency the predictions of quantum mechanics must be independent of who is supposed to make the first measurement and to collapse the system's state. It is easy to see from the discussion above that this condition is satisfied.

What worried EPR was that after Alice's measurement there is a direction in which Bob will never find $+\frac{1}{2}$ for the positron's spin, and this direction depends on what direction Alice chooses to use. This fact seems to imply that the positron somehow 'knows' what Alice measured for the electron, and the collapse of the system's state from (6.33) to (6.34) seems to confirm this suspicion. Since relativity forbids news of Alice's work on the electron from influencing the positron at the time of Bob's measurement, EPR argued that the required information must have travelled out with the positron in the form of a hidden variable which was correlated at the time of the nuclear decay with a matching hidden variable in the electron.

The existence of hidden variables would explain the probabilistic nature of quantum mechanics (which Einstein intensely disliked) because the uncertain outcomes of experiments would reflect our ignorance of the values taken by the hidden variables; the uncertainty would be banished once a better theory gave us access to these variables.

Bell's inequality Remarkably, Bell was able to show that any hidden variable theory will yield a weaker correlation than quantum mechanics between the measurements of Alice and Bob as functions of the angle θ between their chosen directions. Let's denote the results of Alice's and Bob's measurements by $\sigma_{\rm A} = \pm\frac{1}{2}$ and $\sigma_{\rm B} = \pm\frac{1}{2}$ and calculate the expectation value of the product $\sigma_{\rm A}\sigma_{\rm B}$. There are just four cases to consider, so the desired expectation value is

$$\begin{aligned}\langle\sigma_{\rm A}\sigma_{\rm B}\rangle = \tfrac{1}{4}\{&P_{\rm A}(+)P_{\rm B}(+|{\rm A}+) + P_{\rm A}(-)P_{\rm B}(-|{\rm A}-)\\ &- P_{\rm A}(+)P_{\rm B}(-|{\rm A}+) - P_{\rm A}(-)P_{\rm B}(+|{\rm A}-)\},\end{aligned}\tag{6.38}$$

where $P_{\rm A}(+)$ is the probability that Alice obtains $\sigma_{\rm A} = +\frac{1}{2}$ and $P_{\rm B}(-|{\rm A}+)$ is the probability that Bob finds $\sigma_{\rm B} = -\frac{1}{2}$ given that Alice has measured $\sigma_{\rm A} = +\frac{1}{2}$. Since nothing is known about the orientation of the electron before Alice makes her measurement

$$P_{\rm A}(+) = P_{\rm A}(-) = \tfrac{1}{2}.\tag{6.39}$$

We showed above (eq. 6.37) that $P_{\rm B}(+|{\rm A}+) = \sin^2(\theta/2)$, so

$$P_{\rm B}(-|{\rm A}+) = 1 - P_{\rm B}(+|{\rm A}+) = \cos^2(\theta/2).\tag{6.40}$$

Putting these results into equation (6.38) we have

$$\langle\sigma_{\rm A}\sigma_{\rm B}\rangle = \tfrac{1}{4}\{\sin^2(\theta/2) - \cos^2(\theta/2)\} = -\tfrac{1}{4}\cos\theta = -\tfrac{1}{4}\mathbf{a}\cdot\mathbf{b},\tag{6.41}$$

which agrees with Alice's simple argument when $\mathbf{a} \simeq \pm\mathbf{b}$.

Consider now the case that the result of measuring the electron's spin in the direction $\mathbf{a}$ is completely determined by the values taken by hidden variables in addition to $\mathbf{a}$. That is, if we knew the values of these variables, we could predict with certainty the result of measuring the component of the electron's spin in the direction of any unit vector $\mathbf{a}$ and Alice is only uncertain what result she will get because she is ignorant of the values of the hidden variables. We consider the variables to be the components of some n-dimensional vector $\mathbf{v}$, and have that the result of measuring the electron's spin along $\mathbf{a}$ is a function $\sigma_{\rm e}(\mathbf{v},\mathbf{a})$ that takes the values $\pm\frac{1}{2}$ only. Similarly, the result of measuring the positron's spin along a unit vector $\mathbf{b}$ is a function $\sigma_{\rm p}(\mathbf{v},\mathbf{b})$ that is likewise restricted to the values $\pm\frac{1}{2}$. As Alice argued, conservation of angular momentum implies that

$$\sigma_{\rm e}(\mathbf{v},\mathbf{a}) = -\sigma_{\rm p}(\mathbf{v},\mathbf{a}).\tag{6.42}$$

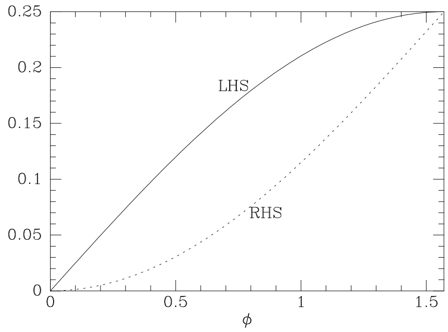


Figure 6.1 For a family of choices of the vectors **a**, **b** and **c**, quantum mechanics predicts that the left side of Bell's inequality (6.46) is larger than the right side, contrary to the prediction of any hidden-variable theory.

The outcome of a measurement is uncertain because the value of **v** is uncertain. We quantify whatever knowledge we do have by assigning a probability density $\rho(\mathbf{v})$ to **v**, which is such that the probability that **v** lies in the infinitesimal n-dimensional volume $\mathrm{d}^n\mathbf{v}$ is $\mathrm{d}P = \rho(\mathbf{v})\,\mathrm{d}^n\mathbf{v}$. In terms of ρ the expectation value of interest is

$$\begin{aligned}\langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{b})\rangle &= \int \mathrm{d}^n\mathbf{v}\,\rho(\mathbf{v})\sigma_\mathrm{e}(\mathbf{v},\mathbf{a})\sigma_\mathrm{p}(\mathbf{v},\mathbf{b})\\ &= -\int \mathrm{d}^n\mathbf{v}\,\rho(\mathbf{v})\sigma_\mathrm{e}(\mathbf{v},\mathbf{a})\sigma_\mathrm{e}(\mathbf{v},\mathbf{b}),\end{aligned} \tag{6.43}$$

where the second equality uses equation (6.42).

If **c** is a third unit vector, the fact that $\sigma_\mathrm{e}^2(\mathbf{v},\mathbf{b}) = \frac{1}{4}$ allows us to write

$$\begin{aligned}\langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{b})\rangle - \langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{c})\rangle &= -\int \mathrm{d}^n\mathbf{v}\,\rho(\mathbf{v})\sigma_\mathrm{e}(\mathbf{v},\mathbf{a})\{\sigma_\mathrm{e}(\mathbf{v},\mathbf{b}) - \sigma_\mathrm{e}(\mathbf{v},\mathbf{c})\}\\ &= -\int \mathrm{d}^n\mathbf{v}\,\rho(\mathbf{v})\sigma_\mathrm{e}(\mathbf{v},\mathbf{a})\sigma_\mathrm{e}(\mathbf{v},\mathbf{b})\{1 - 4\sigma_\mathrm{e}(\mathbf{v},\mathbf{b})\sigma_\mathrm{e}(\mathbf{v},\mathbf{c})\}.\end{aligned} \tag{6.44}$$

We now take the absolute value of each side and note that the curly bracket in the integral is non-negative, while the product $\sigma_\mathrm{e}(\mathbf{v},\mathbf{a})\sigma_\mathrm{e}(\mathbf{v},\mathbf{b})$ in front of it fluctuates between $\pm\frac{1}{4}$. Hence we obtain an upper limit on the value of the integral by replacing $\sigma_\mathrm{e}(\mathbf{v},\mathbf{a})\sigma_\mathrm{e}(\mathbf{v},\mathbf{b})$ by $\frac{1}{4}$, and have

$$|\langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{b})\rangle - \langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{c})\rangle| \le \tfrac{1}{4}\int \mathrm{d}^n\mathbf{v}\,\rho(\mathbf{v})\{1 - 4\sigma_\mathrm{e}(\mathbf{v},\mathbf{b})\sigma_\mathrm{e}(\mathbf{v},\mathbf{c})\}. \tag{6.45}$$

We break the right side into two integrals. The first, $\int \mathrm{d}^n\mathbf{v}\,\rho(\mathbf{v})$ evaluates to unity because ρ is a probability density, while changing $\mathbf{b} \to \mathbf{c}$ and $\mathbf{a} \to \mathbf{b}$ in equation (6.43) we see that the the second integral evaluates to $-4\langle\sigma_\mathrm{e}(\mathbf{b})\sigma_\mathrm{p}(\mathbf{c})\rangle$. Hence we have that

$$|\langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{b})\rangle - \langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{c})\rangle| \le \tfrac{1}{4} + \langle\sigma_\mathrm{e}(\mathbf{b})\sigma_\mathrm{p}(\mathbf{c})\rangle. \tag{6.46}$$

This is **Bell's inequality**, which must hold for any three unit vectors **a**, **b** and **c** if hidden variables exist.

An obvious question is whether Bell's inequality is consistent with the quantum-mechanical result $\langle\sigma_\mathrm{e}(\mathbf{a})\sigma_\mathrm{p}(\mathbf{b})\rangle = -\frac{1}{4}\mathbf{a}\cdot\mathbf{b}$ (eq. 6.41). When we substitute this expression into each side we get

$$\text{LHS} = \tfrac{1}{4}|\mathbf{a}\cdot(\mathbf{b}-\mathbf{c})| \quad ; \quad \text{RHS} = \tfrac{1}{4}(1 - \mathbf{b}\cdot\mathbf{c}). \tag{6.47}$$

Let's choose $\mathbf{a}\cdot\mathbf{b} = 0$ and $\mathbf{c} = \mathbf{b}\cos\phi + \mathbf{a}\sin\phi$ so as we increase the parameter ϕ from zero to $\pi/2$ **c** swings continuously from **b** to **a**. For this choice of **c** we easily find that

$$\text{LHS} = \tfrac{1}{4}|\sin\phi| \quad ; \quad \text{RHS} = \tfrac{1}{4}(1 - \cos\phi). \tag{6.48}$$

These expressions for the left and right sides of Bell's inequality are plotted in Figure 6.1: we see that the inequality is violated for all values of ϕ other than 0 and $\pi/2$. Thus the quantum-mechanical result is inconsistent with Bell's inequality and is therefore inconsistent with the existence of hidden variables.

Inequalities similar to (6.46) can be derived for systems other than spin-half particles, including pairs of entangled photons. Experiments with photons have produced results that agree with the predictions of quantum mechanics to sufficient precision that they violate the relevant Bell inequalities.[4] Consequently, these experiments rule out the possibility that hidden variables exist.

What general conclusions can we draw from the EPR experiment?

- A measurement *both* updates our knowledge of a system *and* disturbs the system. Alice's measurement disturbs the electron but not the positron, and gains her information about both particles.
- Quantum mechanics requires wholistic thinking: when studying the EPR experiment we must consider the system formed by both particles together rather than treating the particles in isolation. We shall encounter a more spectacular example of this requirement below in connection with ideal gases.
- Many discussions of the EPR experiment generate needless confusion by supposing that after Alice has measured $+\frac{1}{2}$ for the component of the electron's spin parallel to $\mathbf{a}$, the spin is aligned with $\mathbf{a}$. We shall see in §7.4.2 that the electron also has half a unit of angular momentum in each of the x and y directions, although the signs of these other components are unknown when we know the value of s_z. Hence the most Alice can know about the orientation of the spin vector is that it lies in a particular hemisphere. Whatever hemisphere Alice determines, she can argue that the positron's spin lies in the opposite hemisphere. So if Alice finds the electron's spin to lie in the northern hemisphere, she concludes that the positron's spin lies in the southern hemisphere. This knowledge excludes only one result from the myriad of possibilities open to Bob: namely he cannot find $s_z = +\frac{1}{2}$. He is unlikely to find $+\frac{1}{2}$ if he measures the component of spin along a vector $\mathbf{b}$ that lies close to the z axis because the hemisphere associated with this result has a small overlap with the southern hemisphere, but since there is an overlap, the result $+\frac{1}{2}$ is not excluded. Contrary to the claims of EPR, the results of Bob's measurements are consistent with the hemisphere containing the positron's spin being fixed at the outset and being unaffected by Alice's measurement.
- The experimental demonstration that Bell inequalities are violated establishes that quantum mechanics will not be superseded by a theory in which the spin vector has a definite direction. In §7.4.1 we shall see that macroscopic objects only appear to have well defined orientations because they are not in states of well-defined spin. That is, the idea that a vector points in a well defined direction is a classical notion and not applicable to objects such as electrons that do have a definite spin. This idea is an old friend from which we part company as sadly as after studying relativity we parted company with the concept of universal time. The world we grew accustomed to in playgroup is not the real world, but an approximation to it that is useful on macroscopic scales. The study of physics forces one to move on and let childish things go.

---

[4] e.g., W. Tittel et al., PRL, **81**, 3563 (1998)

6.2 Quantum computing

There's an old story about a mathematician at the court of the Chinese emperor. The mathematician had advised the emperor wisely and the emperor, wishing to express his gratitude in a manner worthy of his greatness, asked the mathematician to name the reward he would like to receive. "Oh great Emperor, your offer is too liberal for one who has rendered you such a slight service. Let a chess board be brought and one grain of rice be placed on the first square, two on the second, four on the third, eight on the fourth, and so on till every square of the board has received an allocation of rice." The emperor was pleased by the modesty of the mathematician's proposal and ordered it be done. Great was his shock and annoyance the next day when it was reported to him that all the rice in his great silos had proved insufficient to pay the mathematician his due. For $2^{64}-1 \sim 10^{19}$ grains of rice would be needed to supply the 64 squares on the board. That's $\sim 10^{12}$ tons of rice and vastly more than all the rice on the planet.[5]

What is the relevance of this old story for quantum mechanics? We have seen that a system made of two two-state systems has four basis states. If we add a further two-state system to this four-state composite system, we obtain a system with $2 \times 4 = 8$ basis states. By the time we have built a system from 64 two-state systems, our composite system will have $2^{64} \sim 10^{19}$ basis states. Sixty four two-state systems might be constructed from 64 atoms or even 64 electrons, so could be physically miniscule. But to calculate the dynamics of this miniscule system we would have to integrate the equations of motion of 10^{19} amplitudes! This is *seriously* bad news for physics.

The idea behind quantum computing is to turn this disappointment for physics into a boon for mathematics. *We* may not be able to solve 10^{19} equations of motion, but Nature can evolve the physical system, and appropriate measurements made on the system should enable us to discover what the results of our computations would have been if we had the time to carry them out. If this approach to computation can be made to work in practice, calculations will become possible that could never be completed on a conventional computer.

The first step towards understanding how a quantum computer would work is to map integers onto the basis states of our system. In this context we refer to a two-state system as a **qubit** and call its basis states $|0\rangle$ and $|1\rangle$. A set of N qubits forms a **register**, which has a complete set of states of the form $|x\rangle|x'\rangle\cdots|x''\rangle$, where x, x', etc., $=0,1$ indicate the states of the constituent qubits. Now given a number in binary form, such as $7 = 4+2+1 = 111$, we associate it with the basis state of the register $|0\rangle\ldots|0\rangle|1\rangle|1\rangle|1\rangle$. In this way we establish a one to one correspondence between the integers 0 to $2^N - 1$ and the basis states of a register that comprises N qubits. We use this correspondence to establish a more compact notation for the basis states of the register, writing $|7\rangle$ instead of $|0\rangle\ldots|0\rangle|1\rangle|1\rangle|1\rangle$, etc.

This arrangement mirrors the correspondence in a classical computer between numbers and the states of a classical register formed by N classical two-state systems or **bits**. The crucial difference between quantum and classical registers is that whereas a classical register is always in a state that is associated with a definite number, the generic state $|\psi\rangle$ of a quantum register is a linear combination of states that are associated with different numbers:

$$|\psi\rangle = \sum_{j=0}^{2^N-1} c_j|j\rangle. \tag{6.49}$$

Thus nearly all states of a quantum register are *not* associated with individual numbers but with all representable numbers simultaneously. We shall see

---

[5] According to the International Rice Search Institute, in 2007 global rice production was 650 million tons.

that this ability of a single state of a quantum register to be associated with a huge number of integers enables a quantum computer to conduct massively parallel computations.

The central processor unit (**CPU**) of a classical computer is a programmable mechanism that reads a number n from an input register and places the number $f(n)$ into the output register, where f is the function that the CPU is currently programmed to evaluate. By analogy one might imagine that a quantum computer would consist of a quantum register and a programmable Hamiltonian H that would cause the state $|n\rangle$ to evolve in some specified time T into the state $|f(n)\rangle = \mathrm{e}^{-\mathrm{i}HT/\hbar}|n\rangle$. Unfortunately this conception is flawed because this machine could not evaluate any function that took the same value on different arguments, so $f(n) = f(m) = F$, say, for some values $n \neq m$. To see why the computer could not evaluate such a function recall that the operator $U \equiv \mathrm{e}^{-\mathrm{i}HT/\hbar}$ is unitary, so it has an inverse $U^\dagger$. But we have $U|n\rangle = U|m\rangle = |F\rangle$, and if we apply $U^\dagger$ to $|F\rangle$ we must get *both* $|m\rangle$ and $|n\rangle$, which is absurd.

We get around this problem by making our quantum computer slightly more complex: we let it have *two* registers, a **control register** X and a **data register** Y. The computer then has a basis of states $|x\rangle|y\rangle$, where x is the number stored in the control register and y is the number stored in the data register. We conjecture that we can find a Hamiltonian such that for any function f the state $|x\rangle|y\rangle$ evolves in time T into the state $|x\rangle|y + f(x)\rangle$. Adding the second register solves the problem we encountered above because applying U to $|n\rangle|y\rangle$ we get $|n\rangle|y+F\rangle$ which is a different state from what we get when we apply U to $|m\rangle|y\rangle$, namely $|m\rangle|y+F\rangle$: adding the extra register allows the computer to remember the state it was in before the machine cycle started, and this memory makes it logically possible for $U^\dagger$ to restore the earlier state.

Adding the second register may have demolished an objection to our original most naive proposal, but is it really possible to construct a time-evolution operator that would enable us to evaluate any function $f(x)$? This question is answered affirmatively in two stages. First one defines a handful of unitary operators U that perform basic bit manipulations on our registers, and shows that using a sequence of such operators one can perform any of the standard arithmetical operations, adding, subtracting, multiplying and dividing. Second, for each of these operators U one designs an experiment in which U gives the evolution of a two-state quantum system over some time interval. Currently many groups use photons as qubits, identifying $|0\rangle$ and $|1\rangle$ with either right- and left-handed circular polarisation, or with linear polarisation in two orthogonal directions. Other groups use electrons as qubits, identifying $|0\rangle$ and $|1\rangle$ as states in which the spin in some given direction is either $\frac{1}{2}$ or $-\frac{1}{2}$. All such work with real qubits is extremely challenging and in its infancy, but it has already established that there is no objection in principle to realising the simple unitary operators that quantum computing requires. It is too early to tell what physical form qubits will take when quantum computing becomes a mature technology. Consequently, we leave to one side the question of how our operators are to be realised and focus instead on what operators we require and what could be achieved with them when they have been realised.

The simplest computer has two one-qubit registers, with a basis of states $|0\rangle|0\rangle$, $|0\rangle|1\rangle$, $|1\rangle|0\rangle$ and $|1\rangle|1\rangle$ – we shall refer to basis states of a register with any number of qubits ordered thus by increasing value of the stored number as the **computational basis**. In the computational basis of our two-qubit system, the operator U_+ that performs addition ($|x\rangle|y\rangle \rightarrow |x\rangle|y+x\rangle$) has the unitary matrix[6]

$$U_+ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{6.50}$$

To justify this claim we note that

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \delta \\ \gamma \end{pmatrix} \tag{6.51}$$

so U_+ causes the state of the computer

$$|\psi\rangle = \alpha|0\rangle|0\rangle + \beta|0\rangle|1\rangle + \gamma|1\rangle|0\rangle + \delta|1\rangle|1\rangle \tag{6.52}$$

to evolve into

$$U_+|\psi\rangle = \alpha|0\rangle|0\rangle + \beta|0\rangle|1\rangle + \gamma|1\rangle|1\rangle + \delta|1\rangle|0\rangle, \tag{6.53}$$

so the second qubit is indeed incremented by the first modulo 2.

U_+ is a simple example of an operator in which the state of the data register is changed in a way that depends on the state of the control register while the state of the control register stays the same. Such operators are called **controlled-U operators**. Another useful operator is the controlled-phase operator, which in the computational basis has the matrix

$$U_\phi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \mathrm{e}^{\mathrm{i}\phi} \end{pmatrix}. \tag{6.54}$$

U_ϕ has no effect on the first three states of the computational basis, and it multiplies the phase of the last state by $\mathrm{e}^{\mathrm{i}\phi}$. It is straightforward to show that

$$U_\phi|x\rangle|y\rangle = \mathrm{e}^{\mathrm{i}xy\phi}|x\rangle|y\rangle \tag{6.55}$$

by checking that the two sides match for all four possible values of (x, y).

It can be shown that *any* unitary transformation of an n-qubit register can be simulated if we augment U_+ and U_ϕ with two operators that work on just one qubit. One of these extra operators is the **phase operator** U^1_ϕ, which leaves $|0\rangle$ invariant and increments the phase of $|1\rangle$ by ϕ:

$$\left.\begin{array}{l} U^1_\phi|0\rangle = |0\rangle \\ U^1_\phi|1\rangle = \mathrm{e}^{\mathrm{i}\phi}|1\rangle \end{array}\right\} \quad \Leftrightarrow \quad U^1_\phi|x\rangle = \mathrm{e}^{\mathrm{i}\phi x}|x\rangle \quad \Leftrightarrow \quad U^1_\phi = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{e}^{\mathrm{i}\phi} \end{pmatrix}. \tag{6.56}$$

The other single-qubit operator that we need is the **Hadamard operator**, which in the computational basis, $|0\rangle$ $|1\rangle$, has the matrix

$$U_\mathrm{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{6.57}$$

The Hadamard operator takes a state that represents a number, such as $|0\rangle$, and turns it into a state that is a linear combination of the two representable numbers: $U_\mathrm{H}|0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Conversely, because $U_\mathrm{H}^2 = I$ so U_H is

---

[6] Here $x + y$ must be understood to mean $x + y$ mod 2 because quantum computers like classical computers do arithmetic modulo one more than the largest number that they can store.

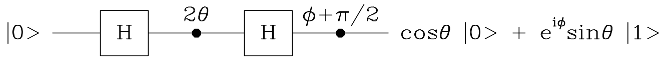


Figure 6.2 Schematic diagram to show how two Hadamard operators and two phase shift operators suffice to transform $|0\rangle$ into an arbitrary state of a qubit.

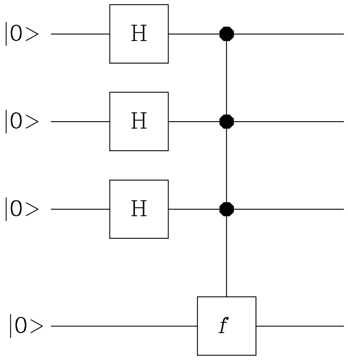


Figure 6.3 Evaluating f on every argument simultaneously. The top three qubits form the control register, which is initially in the state $|0\rangle$.

its own inverse, it turns these linear combinations of numbers into actual numbers: $U_{\rm H}(|0\rangle+|1\rangle)/\sqrt{2}=|0\rangle$.

Complex operations on qubits can be built up by sequences of phase and Hadamard operators and such sequences are conveniently described using the graphical notation of Figure 6.2. Each qubit is represented by a line along which the state of the qubit flows from left to right. In the simple example shown, the state $|0\rangle$ is converted by the first Hadamard operator to $(|0\rangle+|1\rangle)/\sqrt{2}$, and $U^1_{2\theta}$ converts this to

$$\frac{1}{\sqrt{2}}\left(|0\rangle+{\rm e}^{2{\rm i}\theta}|1\rangle\right). \tag{6.58a}$$

After the next Hadamard operator this becomes

$$\begin{aligned}\tfrac{1}{2}\left(|0\rangle+|1\rangle+{\rm e}^{2{\rm i}\theta}(|0\rangle-|1\rangle)\right)&=\tfrac{1}{2}\left\{\left(1+{\rm e}^{2{\rm i}\theta}\right)|0\rangle+\left(1-{\rm e}^{2{\rm i}\theta}\right)|1\rangle\right\}\\&={\rm e}^{{\rm i}\theta}\left\{\cos\theta|0\rangle-{\rm i}\sin\theta|1\rangle\right\}.\end{aligned} \tag{6.58b}$$

Finally, application of the phase-shift operator $U^1_{\phi+\pi/2}$ converts this to

$$|\psi\rangle\equiv{\rm e}^{{\rm i}\theta}\left(\cos\theta|0\rangle+{\rm e}^{{\rm i}\phi}\sin\theta|1\rangle\right). \tag{6.58c}$$

By choosing the values of θ and ϕ appropriately, we can make $|\psi\rangle$ any chosen state of the qubit. Thus the phase-shift and Hadamard operators form a complete set of single-qubit operators.

If we apply a Hadamard operator to each qubit of an 2-qubit register that is initially in the state $|0\rangle|0\rangle$, we get

$$\begin{aligned}(U_{\rm H}|0\rangle)(U_{\rm H}|0\rangle)&=\tfrac{1}{2}(|0\rangle+|1\rangle)(|0\rangle+|1\rangle)\\&=\tfrac{1}{2}(|1\rangle|1\rangle+|1\rangle|0\rangle+|0\rangle|1\rangle+|0\rangle|0\rangle)\\&=\tfrac{1}{2}(|3\rangle+|2\rangle+|1\rangle+|0\rangle).\end{aligned} \tag{6.59}$$

That is, by setting the register to zero and then applying a Hadamard operator to each of its qubits, we put the register into a linear superposition of the states associated with each representable number. It is easy to see that this result generalises to n-qubit registers.[7] Using this trick we can simultaneously evaluate a function on every representable argument, simply by evaluating the function on the state of the control register immediately after it has been processed by the Hadamard operators. Figure 6.3 illustrates this process, which is described by the equations

---

[7] In fact, applying Hadamard operators to the qubits of an n-qubit register when it is set to *any* number will put the register into a linear superposition of states associated with all representable numbers, but if the initial state of the register differs from $|0\rangle$, exactly half of the coefficients in the sum will be $-2^{-n/2}$ and half $+2^{-n/2}$ (Problem 6.5).

Box 6.2: Deutsch's algorithm

Given a function $f(x)$ that takes an n-bit argument and returns either 0 or 1, the exercise is to determine whether f is a constant or 'balanced' function. To this end we build a computer with an n-qubit control register and a single qubit data register. We set the control register to $|0\rangle$ and the data register to $|1\rangle$ and operate on every qubit with the Hadamard operator $U_{\rm H}$. Then the computer's state is

$$\frac{1}{2^{(n+1)/2}}\left(\sum_{x=0}^{2^n-1}|x\rangle\right)(|0\rangle-|1\rangle). \tag{1}$$

Now we evaluate the function f in the usual way, after which the computer's state is

$$\frac{1}{2^{(n+1)/2}}\left(\sum_x |x\rangle(|f(x)\rangle-|1+f(x)\rangle)\right). \tag{2}$$

Given that $f(x)=0,1$, it is straightforward to convince oneself that $(|f(x)\rangle-|1+f(x)\rangle)=(-1)^{f(x)}(|0\rangle-|1\rangle)$ so the computer's state can be written

$$\frac{1}{2^{(n+1)/2}}\left(\sum_x(-1)^{f(x)}|x\rangle\right)(|0\rangle-|1\rangle). \tag{3}$$

We now operate on every qubit with $U_{\rm H}$ for a second time. The data register returns to $|1\rangle$ because $U_{\rm H}$ is its own inverse, while the control register only returns to $|0\rangle$ if we can take the factor $(-1)^{f(x)}$ out of the sum over x, making the state of the control register a multiple of $\sum_x|x\rangle$; if f is 'balanced', half of the factors $(-1)^{f(x)}$ are $+1$ and half -1 and in this case $U_{\rm H}$ moves the control register to a state $|y\rangle$ for $y\neq 0$ (Problem 6.5). Hence by measuring the state of the control register, we discover whether f is constant or balanced: if the control register is set to zero, f is constant, and if it holds any other number, f is balanced.

$$|0\rangle|0\rangle\xrightarrow{U_{\rm H}}\frac{1}{2^{n/2}}\sum_{x=0}^{2^n-1}|x\rangle|0\rangle\xrightarrow{f}\frac{1}{2^{n/2}}\sum_{x=0}^{2^n-1}|x\rangle|f(x)\rangle. \tag{6.60}$$

After the evaluation of f, the computer's state depends on every possible value of f. So the state of a 64-qubit computer will depend on the $2^{64}\simeq 10^{19}$ possible values of f. By exploiting this fact, can we conduct massively parallel computations with just a pair of quantum registers?

The question is, how can we learn about the values that f takes? An obvious strategy is to read off a numerical value X from the control register by collapsing each of its qubits into either the state $|0\rangle$ or the state $|1\rangle$. Once this has been done, the state of the composite system $|x\rangle|y\rangle$ will have collapsed from that given on the right of (6.60) to $|X\rangle|f(X)\rangle$, so $f(X)$ can be determined by inspecting each of the qubits of the data register. The trouble with this strategy is that it only returns one value of f, and that for a random argument X. Hence if our quantum computer is to outperform a classical computer, we must avoid collapsing the computer's state by reading its registers. Instead we should try to answer questions about f that have simple answers but ones that involve all the values taken by f.

For example, suppose we know that $f(x)$ only takes the values 0 and 1, and that it is either a constant function (i.e., either $f(x)=1$ for all x, or $f(x)=0$ for all x) or it is a 'balanced function' in the sense that $f(x)=0$ for half of the possible values of x and 1 for the remaining values. The question we have to answer is "is f constant or balanced?" With a classical computer you would have to keep evaluating f on different values of x until either you got two different values (which would establish that f was balanced) or more than half of the possible values of x had been tried (which would establish that f was constant). In Box 6.2 we show that from (6.60) we can discover whether f is constant or balanced with only a handful of machine cycles.

The algorithm given in Box 6.2 is an extension of one invented by Deutsch[8], which was an early example of how the parallel-computing potential of a quantum computer could be harnessed. Subsequently algorithms were developed that dramatically accelerate database searches[9] and the decomposition of large numbers into their prime factors. The usefulness of the internet depends on effective cryptography, which currently relies on the difficulty of prime-number decomposition. Hence by rendering existing cryptographic systems ineffective, the successful construction of a quantum computer would have a big impact on the world economy.

Notwithstanding strenuous efforts around the world, quantum computing remains a dream that will not be realised very soon. Its central idea is that the the integers up to $2^N - 1$ can be mapped into the base states of an N-qubit quantum register, so a general state of such a register is associated with all representable integers, and the time evolution of the register involves massively parallel computing. The field is challenging both experimentally and theoretically. The challenge for theorists is to devise algorithms that extract information from a quantum register given that any measurement of the register collapses its state and thus erases much of the information that was encoded in it before a measurement was made. Experimentally, the challenge is to isolate quantum registers from their environment sufficiently well that they do not become significantly entangled with the environment during a computation. We discuss the process of becoming entangled with the environment in the next section.

6.3 The density operator

To this point in this book we have assumed that we know what quantum state our system is in. For macroscopic objects this assumption is completely unrealistic, for how can we possibly discover the quantum states of the $\sim 10^{23}$ carbon atoms in a diamond, or even the $\sim 10^5$ atoms in a protein molecule? To achieve this goal for a diamond, at least 10^{23} observables would have to be measured, and the number would in reality be vastly greater because individual atoms would be entangled with one another, making the state of the diamond a linear combination of basis states of the form $|a_1\rangle|a_2\rangle\dots|a_N\rangle$, where $|a_i\rangle$ denotes a state of the i^{th} atom. It is time we squared up to the reality of our ignorance of the quantum states of macro- and meso-scopic objects.

Actually, we need to be cautious even when asserting that we know the quantum state of atomic-scale objects. The claim that the state of a system is known is generally justified by the assertion that a measurement has just been made, with the result that the system's state has been collapsed into a known eigenstate of the operator of the given observable. This procedure for establishing the quantum state of a system is unrealistic in that it makes no allowance for experimental error, which we all know to be endemic in real laboratories: real experiments lead to the conclusion that the value of an observable is $x \pm y$, which is shorthand for "the probability distribution for the value of the observable is centred on x and has a width of the order y." Since the measurement leaves the value of the observable uncertain, it does not determine the quantum state precisely either.

---

[8] D. Deutsch, *Proc. R. Soc.*, **400**, 97 (1985)

[9] L. K. Grover, *STOC'96*, 212 (1996)

Let us admit that we don't know what state our system is in, but conjecture that the system is in *one* of a complete set of states $\{|n\rangle\}$, and for each value of n assign a probability p_n that it's in the state $|n\rangle$. It's important to be clear that we are *not* saying that the system is in the state

$$|\phi\rangle = \sum_n p_n |n\rangle. \tag{6.61}$$

That is a well-defined quantum state, and we are admitting that we don't know the system's state. What we are saying is that the system may be in state $|n\rangle$, or in state $|n+1\rangle$, or state $|n+2\rangle$, and assigning probabilities p_n, $p_{n+1}, \ldots$ to each of these possibilities.

Given this incomplete information, the expectation value of measuring some observable Q will be p_n times the expectation value that Q will have if the system is in the state $|n\rangle$, plus p_{n+1} times the expectation value for the case that the system is in the state $|n+1\rangle$, etc. That is

$$\overline{Q} = \sum_n p_n \langle n|Q|n\rangle, \tag{6.62}$$

where we have introduced a new notation $\overline{Q}$ to denote the expectation value of Q when we have incomplete knowledge. When our knowledge of a system is incomplete, we say that the system is in an **impure state**, and correspondingly we sometimes refer to a regular state $|\psi\rangle$ as a **pure state**. From the present point of view this terminology is unfortunate for a system 'in an impure state' may have a perfectly well-defined state $|\psi\rangle$ even if we are uncertain what it is.

It is instructive to rewrite equation (6.62) by inserting either side of Q identity operators $I = \sum_j |q_j\rangle\langle q_j|$ that are made out of the eigenkets of Q. Then we have

$$\overline{Q} = \sum_{nkj} p_n \langle n|q_k\rangle\langle q_k|Q|q_j\rangle\langle q_j|n\rangle = \sum_{nj} q_j p_n |\langle q_j|n\rangle|^2, \tag{6.63}$$

where the second equality follows from $Q|q_j\rangle = q_j|q_j\rangle$ and the orthonormality of the kets $|q_j\rangle$. Equation (6.63) states that the expectation value of Q is the sum of the possible measurement values q_j times the probability $p_n|\langle q_j|n\rangle|^2$ of obtaining this value, which is the product of the probability of the system being in the state $|n\rangle$ and the probability of obtaining q_j in the case that it is.

Now consider the **density operator**

$$\rho \equiv \sum_n p_n |n\rangle\langle n|, \tag{6.64}$$

where the p_n are the probabilities introduced above. This definition is reminiscent of the definition

$$Q = \sum_j q_j |q_j\rangle\langle q_j| \tag{6.65}$$

of the operator associated with an observable (eq. 2.9). In particular, ρ is a Hermitian operator because the p_n are real. It should not be considered an observable, however, because the p_n are subjective not objective: they quantify our state of knowledge rather than hard physical reality. For example, if our records of the results of measurements become scrambled, perhaps through some failure of electronics in the data-acquisition system, our values of the p_n will change but the system will not. By contrast the spectrum $\{q_j\}$

Box 6.3: Properties of Tr

The trace operator Tr extracts a complex number from an operator. We now show that although its definition (6.67) is in terms of a particular basis $\{|m\rangle\}$, its value is independent of the basis used. Let $\{|q_j\rangle\}$ be any other basis. Then we insert identity operators $I=\sum_j |q_j\rangle\langle q_j|$ either side of A in $\mathrm{Tr}\,A=\sum_n\langle n|A|n\rangle$:

$$\begin{aligned}\mathrm{Tr}\,A&=\sum_{njk}\langle n|q_j\rangle\langle q_j|A|q_k\rangle\langle q_k|n\rangle=\sum_{kjn}\langle q_j|A|q_k\rangle\langle q_k|n\rangle\langle n|q_j\rangle\\&=\sum_j\langle q_j|A|q_j\rangle,\end{aligned}\tag{1}$$

where we have used $I=\sum_n|n\rangle\langle n|$ and $\langle q_k|q_j\rangle=\delta_{kj}$.

Another useful result is that for any two operators A and B, $\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$:

$$\begin{aligned}\mathrm{Tr}(AB)&=\sum_n\langle n|AB|n\rangle=\sum_{nm}\langle n|A|m\rangle\langle m|B|n\rangle\\&=\sum_{nm}\langle m|B|n\rangle\langle n|A|m\rangle=\sum_m\langle m|BA|m\rangle=\mathrm{Tr}(BA).\end{aligned}\tag{2}$$

By making the substitutions $B\to C$ and $A\to AB$ in this result we infer that

$$\mathrm{Tr}(ABC)=\mathrm{Tr}(CAB).\tag{3}$$

of Q is determined by the laws of nature and is independent of the completeness of our knowledge. Thus the density operator introduces a qualitatively new feature into the theory: subjectivity.

To see the point of the density operator, we use equations (6.64) and (6.65) to rewrite the operator product

$$\rho Q=\sum_{nj}p_nq_j|n\rangle\langle n|q_j\rangle\langle q_j|.\tag{6.66}$$

When this equation is premultiplied by $\langle m|$ and postmultiplied by $|m\rangle$ and the result summed over m, the right side becomes the same as the right side of equation (6.63) for $\overline{Q}$. That is,

$$\mathrm{Tr}(\rho Q)\equiv\sum_m\langle m|\rho Q|m\rangle=\overline{Q},\tag{6.67}$$

where 'Tr' is short for 'trace' because the sum over m is of the diagonal elements of the matrix for ρQ in the basis $\{|n\rangle\}$. Box 6.3 derives two important properties of the trace operator.

Equation (6.64) defines the density operator in terms of the basis $\{|n\rangle\}$. What do we get if we express ρ in terms of some other basis $\{|q_j\rangle\}$? To find out we replace $|n\rangle$ by $\sum_j\langle q_j|n\rangle|q_j\rangle$ and obtain

$$\begin{aligned}\rho&=\sum_{njk}p_n\langle q_j|n\rangle\langle n|q_k\rangle\,|q_j\rangle\langle q_k|\\&=\sum_{jk}p_{jk}|q_j\rangle\langle q_k|\qquad\text{where}\quad p_{jk}\equiv\sum_n p_n\langle q_j|n\rangle\langle n|q_k\rangle.\end{aligned}\tag{6.68}$$

This equation shows that whereas ρ is represented by a diagonal matrix in the $\{|n\rangle\}$ basis, in the $\{|q_j\rangle\}$ basis ρ is represented by a non-diagonal matrix. This contrast arises because in writing equation (6.64) we *assumed* that our system was in one of the states of the set $\{|n\rangle\}$, although we were unsure which one. In general if the system is in one of these states, it will definitely not be in *any* of the states $\{|q_j\rangle\}$ because each $|n\rangle$ will be a non-trivial linear combination of kets $|q_j\rangle$. Thus when ρ is expanded in this basis, the expansion does not simply specify a probability to be in each state. Instead it includes complex off-diagonal terms $p_{jk} = \sum_n p_n\langle q_j|n\rangle\langle n|q_k\rangle$ that have no classical interpretation. When we have incomplete knowledge of the state of our system, we will generally not know that the system is in some state of a given complete set, so we should not assume that the off-diagonal elements of ρ vanish. Never the less, we may safely use equation (6.64) because whatever matrix represents ρ in a given basis, ρ is a Hermitian operator and will have a complete set of eigenkets. Equation (6.64) gives the expansion of ρ in terms of its eigenkets. In practical applications we may not know what the eigenkets $|n\rangle$ are, but this need not prevent us using them in calculations.

The importance of ρ is that through equation (6.67) we can obtain from it the expectation value of any observable. As the system evolves, these observables will evolve because ρ evolves. To find its equation of motion, we differentiate equation (6.64) with respect to time and use the TDSE. The differentiation is straightforward because p_n is time-independent: if the system was in the state $|n\rangle$ at time t, at any later time it will certainly be in whatever state $|n\rangle$ evolves into. Hence we have

$$\begin{aligned}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= \sum_n p_n\left(\frac{\partial|n\rangle}{\partial t}\langle n| + |n\rangle\frac{\partial\langle n|}{\partial t}\right)\\ &= \frac{1}{\mathrm{i}\hbar}\sum_n p_n\left(H|n\rangle\langle n| - |n\rangle\langle n|H\right) = \frac{1}{\mathrm{i}\hbar}(H\rho - \rho H).\end{aligned}\tag{6.69}$$

This equation of motion can be written more simply

$$\mathrm{i}\hbar\frac{\mathrm{d}\rho}{\mathrm{d}t} = [H,\rho].\tag{6.70}$$

To obtain the equation of motion of an arbitrary expectation value $\overline{Q} = \mathrm{Tr}(\rho Q)$, we expand the trace in terms of a time-independent basis $\{|a\rangle\}$ and use equation (6.70):

$$\mathrm{i}\hbar\frac{\mathrm{d}\overline{Q}}{\mathrm{d}t} = \mathrm{i}\hbar\sum_a\langle a|\frac{\mathrm{d}\rho}{\mathrm{d}t}Q|a\rangle = \sum_a\langle a|(H\rho - \rho H)Q|a\rangle = \mathrm{Tr}(\rho[Q,H]),\tag{6.71}$$

where the last equality uses equation (3) of Box 6.3. Ehrenfest's theorem (2.34) states that the rate of change of the expectation value Q for a given quantum state is the expectation value of $[Q,H]$ divided by $\mathrm{i}\hbar$, so equation (6.71) states that when the quantum state is uncertain, the expected rate of change of Q is the appropriately weighted average of the rates of change of Q for each of the possible states of the system.

Notice that the density operator and the operators for the Hamiltonian and other observables encapsulate a complete, self-contained theory of dynamics. If we have incomplete knowledge of our system's initial state, use of this theory is mandatory. If we *do* know the initial state, we can still use this apparatus by assigning our system the density operator

$$\rho = |\psi\rangle\langle\psi|\tag{6.72}$$

rather than using the TDSE and extracting amplitudes for possible outcomes of measurements. However, when ρ takes the special form (6.72), the use of the density operator becomes optional (Problem 6.7).

6.3.1 Reduced density operators

We have seen that any physical interaction between two quantum systems is likely to entangle them. No man is an island and no system is truly isolated (except perhaps the entire Universe!) Consequently, a real system is constantly entangling itself with its environment. We now show that even if our system starts in a pure state, once it has entangled itself with its environment, it will be in an impure state.

We consider a system that is comprised of two subsystems: A, which will represent our system, and B, which will represent the environment – the **environment** consists of anything that is dynamically coupled to our system but not observed in sufficient detail for its dynamics to be followed. Let the density operator of the entire system be

$$\rho_{\rm AB} = \sum_{ijkl} |{\rm A};i\rangle|{\rm B};j\rangle\rho_{ijkl}\langle{\rm A};k|\langle{\rm B};l|. \tag{6.73}$$

Let Q be an observable property of subsystem A. The expectation value of Q is

$$\begin{aligned}\overline{Q} &= {\rm Tr}\, Q\rho\\ &= \sum_{mn}\langle{\rm A};m|\langle{\rm B};n|Q\left(\sum_{ijkl}|{\rm A};i\rangle|{\rm B};j\rangle\rho_{ijkl}\langle{\rm A};k|\langle{\rm B};l|\right)|{\rm A};m\rangle|{\rm B};n\rangle\\ &= \sum_{mi}\langle{\rm A};m|Q|{\rm A};i\rangle\sum_n \rho_{inmn},\end{aligned} \tag{6.74}$$

where the second equality exploits the fact that Q operates only on the states of subsystem A, and also uses the orthonormality of the states of each subsystem: $\langle{\rm A};k|{\rm A};m\rangle = \delta_{km}$, etc. We now define the **reduced density operator** of subsystem A to be

$$\rho_{\rm A} \equiv \sum_n \langle{\rm B};n|\rho_{\rm AB}|{\rm B};n\rangle = \sum_{im}|{\rm A};i\rangle\left(\sum_n \rho_{inkn}\right)\langle{\rm A};k|, \tag{6.75}$$

where the second equality uses equation (6.73). In terms of the reduced density operator, equation (6.74) can be written

$$\overline{Q} = \sum_m \langle{\rm A};m|Q\rho_{\rm A}|{\rm A};m\rangle = {\rm Tr}\, Q\rho_{\rm A}. \tag{6.76}$$

Thus the reduced density operator enables us to obtain expectation values of subsystem A's observables without bothering about the states of subsystem B. It is formed from the density operator of the entire system by taking the **partial trace** over the states of subsystem B (eq. 6.75).

Suppose both subsystems start in well-defined states. Then under the TDSE the composite system will evolve through a series of pure states $|\psi,t\rangle$, and at time t the density operator of the composite system will be (cf. 6.72)

$$\rho_{\rm AB} = |\psi,t\rangle\langle\psi,t|. \tag{6.77}$$

If the two subsystems have not become entangled, so $|\psi,t\rangle = |{\rm A},t\rangle|{\rm B},t\rangle$, then the reduced density operator for A is

$$\rho_{\rm A} = |{\rm A},t\rangle\langle{\rm A},t|\sum_i\langle{\rm B};i|{\rm B},t\rangle\langle{\rm B},t|{\rm B};i\rangle = |{\rm A},t\rangle\langle{\rm A},t|, \tag{6.78}$$

where we have used the fact that the set $\{|{\rm B};i\rangle\}$ is a complete set of states for subsystem B. Equation (6.78) shows that so long as the subsystems remain unentangled, the reduced density operator for A has the form expected for a system that is in a pure state. To show that entanglement will generally lead subsystem A into an impure state, we consider the simplest non-trivial example: that in which both subsystems are qubits. Suppose they have evolved into the entangled state

$$|\psi,t\rangle = \frac{1}{\sqrt{2}}\left(|\mathrm{A};0\rangle|\mathrm{B};0\rangle + |\mathrm{A};1\rangle|\mathrm{B};1\rangle\right) \tag{6.79}$$

Then evaluating the trace over the two states of B we find

$$\begin{aligned}\rho_\mathrm{A} &= \tfrac{1}{2}\langle\mathrm{B};0|\left(|\mathrm{A};0\rangle|\mathrm{B};0\rangle + |\mathrm{A};1\rangle|\mathrm{B};1\rangle\right)\left(\langle\mathrm{A};0|\langle\mathrm{B};0| + \langle\mathrm{A};1|\langle\mathrm{B};1|\right)|\mathrm{B};0\rangle \\ &\quad + \tfrac{1}{2}\langle\mathrm{B};1|\left(|\mathrm{A};0\rangle|\mathrm{B};0\rangle + |\mathrm{A};1\rangle|\mathrm{B};1\rangle\right)\left(\langle\mathrm{A};0|\langle\mathrm{B};0| + \langle\mathrm{A};1|\langle\mathrm{B};1|\right)|\mathrm{B};1\rangle \\ &= \tfrac{1}{2}\left(|\mathrm{A};0\rangle\langle\mathrm{A};0| + |\mathrm{A};1\rangle\langle\mathrm{A};1|\right),\end{aligned} \tag{6.80}$$

which is the density operator of a very impure state. Physically this result makes perfect sense: in equation (6.80) ρ_A states that subsystem A has equal probability of being in either $|0\rangle$ or $|1\rangle$, which is consistent with the state (6.79) of the entire system. In that state these two possibilities were associated with distinct predictions about the state of subsystem B, but in passing from ρ_AB to ρ_A we have lost track of these correlations: if we choose to consider system A in isolation, we lose the information carried by these correlations, with the result that we have incomplete information about system A. In this case system A is in an impure state. So long as we recognise that A is part of the larger system AB and we retain the ability to measure both parts of AB, we have complete information, so AB is in a pure state.

In this example system A represents the system under study and system B represents the environment of A, which we defined to be whatever is dynamically coupled to A but incompletely instrumented. If, for example, A is a hydrogen atom, then the electromagnetic field inside the vessel containing the atom would form part of B because a hydrogen atom, being comprised of two moving charged particles, is inevitably coupled to the electromagnetic field. If we start with the atom in its first excited state and the electromagnetic field in its ground state, then atom, field and atom-plus-field are initially all in pure states. After some time the atom-plus-field will evolve into the state

$$|\psi,t\rangle = a_0(t)|\mathrm{A};0\rangle|\mathrm{F};1\rangle + a_1(t)|\mathrm{A};1\rangle|\mathrm{F};0\rangle, \tag{6.81}$$

where $|\mathrm{A};n\rangle$ is the n^th excited state of the atom, while $|\mathrm{F};n\rangle$ is the state of the electromagnetic field when it contains n photons of the frequency associated with transitions between the atom's ground and first-excited states. In equation (6.81), $a_0(t)$ is the amplitude that the atom has decayed to its ground state while $a_1(t)$ is the amplitude that it is still in its excited state. When neither amplitude vanishes, the atom is entangled with the electromagnetic field. If we fail to monitor the electromagnetic field, we have to describe the atom by its reduced density operator

$$\rho_\mathrm{A} = |a_0|^2|\mathrm{A};0\rangle\langle\mathrm{A};0| + |a_1|^2|\mathrm{A};1\rangle\langle\mathrm{A};1|. \tag{6.82}$$

This density operator indicates that the atom is now in an impure state.

In practice a system under study will sooner or later become entangled with its environment, and once it has, we will be obliged to treat the system as one for which we lack complete information. That is, we will have to predict the results of measurements with a non-trivial density operator. The transition of systems in this way from pure states to impure ones is called **quantum decoherence**. Experimental work directed at realising the possibilities offered by quantum computing is very much concerned with arresting the decoherence process by weakening all couplings to the environment.

6.3.2 Shannon entropy

Once we recognise that systems are typically in impure states, it's natural to want to quantify the impurity of a state: for example, if in the definition (6.64) of the density operator, $p_3 = 0.99999999$, then the system is almost certain to be found in the state $|3\rangle$ and predictions made by assuming that the system is in the pure state $|3\rangle$ will not be much in error, while if the largest probability occurring in the sum is 10^{-20}, the effects of impurity will be enormous.

A probability distribution $\{p_i\}$ provides a certain amount of information about the outcome of some investigation. If one probability is close to unity, the information it provides is nearly complete. Conversely, if all the probabilities are small, no outcome is particularly likely and the missing information is large. The question we now address is "what is the appropriate measure of the missing information that remains after a probability distribution $\{p_i\}$ has been specified?"

Logic dictates that the required measure $s(p_1, \ldots, p_n)$ of missing information must have the following properties:

- s must be a continuous, symmetric function of the p_i;
- s should be largest when every outcome is equally likely, i.e., when $p_i = 1/n$ for all i. We define
$$s(\tfrac{1}{n}, \ldots, \tfrac{1}{n}) = s_n \tag{6.83}$$
and require that $s_{n+1} > s_n$ (more possibilities implies more missing information).
- s shall be consistent in the sense that it yields the same missing information when there are different ways of enumerating the possible outcomes of the event.

To grasp the essence of the last requirement, consider an experiment with three possible outcomes x_1, x_2 and x_3 to which we assign probabilities p_1, p_2 and p_3, yielding missing information $s(p_1, p_2, p_3)$. We could group the last two outcomes together into the outcome x_{23}, by which we mean "either x_2 or x_3". Then we assign a probability $p_{23} = p_2 + p_3$ to getting x_{23}, giving missing information $s(p_1, p_{23})$. To this missing information we have to add that associated with resolving the outcome x_{23} into either x_2 or x_3. The probability that we will have to resolve this missing information is p_{23}, and the probability of getting x_2 given that we have x_{23} is p_2/p_{23}, so we argue that
$$s(p_1, p_2, p_3) = s(p_1, p_{23}) + p_{23}s\Big(\frac{p_2}{p_{23}}, \frac{p_3}{p_{23}}\Big). \tag{6.84}$$
This equation is readily generalised: we have n possible outcomes $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$. We gather the outcomes into r groups and let y_1 be the outcome in which one of $x_1, \ldots, x_{k_1}$ was obtained, y_2 the outcome in which one of $x_{k_1+1} \ldots, x_{k_2}$ was obtained etc, and let w_i denote the probability of the outcome y_i. Then since the probability that we get x_1 given that we have already obtained y_1 is p_1/w_1, we have
$$\begin{aligned} s(p_1, \ldots, p_n) = s(w_1, \ldots, w_r) + w_1 s(p_1/w_1, \ldots, p_{k_1}/w_1) + \\ \cdots + w_r s(p_{n-k_r}/w_r, \ldots, p_n/w_r). \end{aligned} \tag{6.85}$$

Since s is a continuous function of its arguments, it suffices to evaluate it for rational values of the arguments. So we assume that there are integers n_i such that $p_i = n_i/N$, where $\sum_i n_i = N$ by the requirement that the probabilities sum to unity. Consider a system in which there are N equally likely outcomes, and from these form n groups, with n_i possibilities in the i^{th} group. Then the probability of the group is p_i and the probability of getting any possibility in the i^{th} group given that the i^{th} group has come up, is $1/n_i$. Hence applying equation (6.85) to the whole system we find
$$s(1/N, \ldots, 1/N) = s(p_1, \ldots, p_n) + \sum_i^n p_i s(1/n_i, \ldots, 1/n_i) \tag{6.86}$$

Box 6.4: Solving Equation (6.88)

Let $s(n) \equiv s_n$. Then equation (6.88) is easily extended to

$$s(mnr\cdots) = s(n) + s(m) + s(r) + \cdots,$$

so with $n = m = r = \cdots$ we conclude that

$$s(n^k) = ks(n).$$

Now let u, v be any two integers bigger than 1. Then for arbitrarily large n we can find m such that

$$\frac{m}{n} \le \frac{\ln v}{\ln u} < \frac{m+1}{n} \quad \Rightarrow \quad u^m \le v^n < u^{m+1}. \tag{1}$$

Since s is monotone increasing,

$$\begin{aligned} s(u^m) \le s(v^n) < s(u^{m+1}) \quad &\Rightarrow \quad ms(u) \le ns(v) < (m+1)s(u) \\ \Rightarrow \quad \frac{m}{n} \le \frac{s(v)}{s(u)} &< \frac{m+1}{n}. \end{aligned} \tag{2}$$

Comparing equation (1) with equation (2), we see that

$$\left|\frac{s(v)}{s(u)} - \frac{\ln v}{\ln u}\right| \le \frac{1}{n} \quad \Rightarrow \quad \left|\frac{s(v)}{\ln v} - \frac{s(u)}{\ln u}\right| \le \epsilon,$$

where $\epsilon = s(u)/(n \ln v)$ is arbitrary small. Thus we have shown that $s(v) \propto \ln v$.

or with the definition (6.83) of s_n,

$$s(p_1, \ldots, p_n) = s_N - \sum_i^n p_i s_{n_i} \qquad \Big(N = \sum_i^n n_i\Big). \tag{6.87}$$

This equation relates s evaluated on a general argument list to the values that s takes when all its arguments are equal. Setting all the $n_i = m$ we obtain a relation that involves only s_n:

$$s_n = s_{nm} - s_m. \tag{6.88}$$

It is easy to check that this functional equation is solved by $s_n = K \ln n$, where K is an arbitrary constant that we can set to unity. In fact, in Box 6.4 it is shown that this is the only monotone solution of equation (6.88). Hence from equation (6.87) we have that the unique measure of missing information is

$$\begin{aligned} s(p_1, \ldots, p_n) &= \ln N - \sum_i^n p_i \ln n_i = -\sum_i p_i (\ln n_i - \ln N) \\ &= -\sum_i p_i \ln p_i. \end{aligned} \tag{6.89}$$

Since every probability p_i is non-negative and less than or equal to one, s is inherently positive. Claude Shannon (1916–2001) first demonstrated[10] that the function (6.89) is the only consistent measure of missing information. Since $s(\mathbf{p})$ turns out to be intimately connected to thermodynamic entropy, it is called the **Shannon entropy** of the probability distribution.

---

[10] C.E. Shannon, *Bell Systems Technical Journal*, **27**, 379 (1948). For a much fuller account, see E.T. Jaynes *Probability Theory: the Logic of Science* Cambridge University Press, 2003.

The Shannon entropy of a density operator ρ is defined to be

$$s(\rho) = -\operatorname{Tr}\rho\ln\rho. \tag{6.90}$$

The right side of this expression involves a function, $\ln(x)$ of the operator ρ. We recall from equation (2.20) that $f(\rho)$ has the same eigenkets as ρ and eigenvalues $f(\lambda_i)$, where λ_i are the eigenvalues of ρ. Hence

$$s = -\operatorname{Tr}(\rho\ln\rho) = -\sum_n \langle n|\sum_i p_i|i\rangle\langle i|\sum_j \ln(p_j)|j\rangle\langle j|n\rangle = -\sum_n p_n\ln p_n. \tag{6.91}$$

Hence s is simply the Shannon entropy of the probability distribution $\{p_i\}$ that appears in the definition (6.64) of ρ.

6.4 Thermodynamics

Thermodynamics is concerned with macroscopic systems about which we don't know very much, certainly vastly less than is required to define a quantum state. For example, the system might consist of a cylinder full of fluid and our knowledge be confined to the chemical nature of the fluid (that it is O_2 or CO_2, or whatever), the mass of fluid, its volume and the temperature of the environment with which it is in equilibrium. In the **canonical picture** we consider that as a result of exchanges of energy with the environment, the energy of the fluid fluctuates around a mean U. The pressure also fluctuates around a mean value P, but the volume $\mathcal{V}$ is well-defined and under our control.

Thermodynamics applies to systems that are more complex than bodies of fluid, for example to a quantity of diamond. In such a case the stress in the material is not fully described by the pressure, and thermodynamic relations involve also the shear stress and the shear strain within the crystal. If the crystal, like quartz, has interesting electrical properties, the thermodynamic relations will involve the electric field within the material and the polarisation that it induces. A fluid is the simplest non-trivial thermodynamic system and therefore the focus of introductory texts, but the principles that it illuminates are of much wider validity. For simplicity we restrict our discussion to fluids.

To obtain relations between the thermodynamic variables from a knowledge of the system's microstructure, we need to assign a probability p_i to each of the system's zillions of quantum states. We argue that the only rational way to assign probabilities to the stationary states of a thermodynamic system is to choose them such that (i) they reproduce any measurements we have of the system, and (ii) they maximise the Shannon entropy. Requirement (ii) follows because in choosing the $\{p_i\}$ we must not specify any information beyond that included when we satisfy requirement (i) – our probabilities must "tell the truth, the whole truth and nothing but the truth". It is straightforward to show (Problem 6.13) that the p_i that maximise the Shannon entropy for given internal energy

$$U \equiv \sum_{\substack{\text{stationary}\\ \text{states } i}} E_i p_i \tag{6.92}$$

are given by

$$p_i = \frac{1}{Z}\mathrm{e}^{-\beta E_i}, \tag{6.93a}$$

where $\beta \equiv 1/(k_\mathrm{B}T)$ is the inverse temperature and

$$Z \equiv \sum_{\substack{\text{stationary}\\ \text{states } i}} \mathrm{e}^{-\beta E_i}. \tag{6.93b}$$

The quantity Z defined above is called the **partition function**; it is manifestly a function of T and less obviously a function of the volume $\mathcal{V}$ and whatever other parameters define the spectrum $\{E_i\}$ of the Hamiltonian. In equation (6.93a) its role is clearly to ensure that the probabilities satisfy the normalisation condition $\sum_i p_i = 1$.

Since the probability distribution (6.93a) maximises the Shannon entropy for given internal energy, we take the density operator of a thermodynamic system to be diagonal in the energy representation and to be given by

$$\rho = \frac{1}{Z} \sum_{\substack{\text{stationary}\\ \text{states } i}} \mathrm{e}^{-\beta E_i} |i\rangle\langle i|. \tag{6.94}$$

This form of the density operator is called the **Gibbs distribution** in honour of J.W. Gibbs (1839–1903), who died before quantum mechanics emerged but had already established that probabilities should given by equation (6.93a).

The sum in equation (6.94) is over quantum states *not* energy levels. It is likely that many energy levels will be highly degenerate and in this case the sum simplifies to $Z = \sum_\alpha g_\alpha \mathrm{e}^{-\beta E_\alpha}$, where α runs over energy levels and g_α is the number of linearly independent quantum states in level α.

The expectation of the Hamiltonian of a thermodynamic system is

$$\overline{H} = \mathrm{Tr}(H\rho) = \sum_n \langle n|H \sum_i p_i |i\rangle\langle i|n\rangle = \sum_n p_n E_n = U, \tag{6.95}$$

where we have used the definition (6.92) of the internal energy. Thus the internal energy U of thermodynamics is simply the expectation value of the system's Hamiltonian. Another important expression for U follows straightforwardly from equations (6.92) and (6.93):

$$U = -\frac{\partial \ln Z}{\partial \beta}. \tag{6.96}$$

We obtain an interesting equation using equation (6.93a) to eliminate the second occurrence of p_n from the extreme right of equation (6.91):

$$s = -\sum_n p_n(-\beta E_n - \ln Z) = \beta U + \ln Z. \tag{6.97}$$

In terms of the **thermodynamic entropy**

$$S \equiv k_{\mathrm{B}} s \tag{6.98}$$

and the **Helmholtz free energy**

$$F \equiv -k_{\mathrm{B}} T \ln Z \tag{6.99}$$

equation (6.97) can be written

$$F = U - TS, \tag{6.100}$$

which in classical thermodynamics is considered to be the definition of the Helmholtz free energy. When we substitute our definition of F into equation (6.96), we obtain

$$U = \frac{\partial(\beta F)}{\partial \beta} = F + \beta\frac{\partial F}{\partial \beta} = F - T\frac{\partial F}{\partial T}. \tag{6.101}$$

Comparing this equation with equation (6.100) we conclude that

$$S = -\frac{\partial F}{\partial T}. \tag{6.102}$$

The difference of equation (6.92) between two similar thermodynamic states is

$$\mathrm{d}U = \sum_i (\mathrm{d}p_i E_i + p_i \mathrm{d}E_i). \tag{6.103}$$

Similarly differencing the definition $S = -k_\mathrm{B} \sum_i p_i \ln p_i$ of the thermodynamic entropy (eqns 6.91 and 6.98), we obtain

$$\mathrm{d}S = -k_\mathrm{B} \sum_i (\ln p_i + 1)\,\mathrm{d}p_i = -k_\mathrm{B} \sum_i \ln p_i \,\mathrm{d}p_i, \tag{6.104}$$

where the second equality exploits the fact that $\{p_i\}$ is a probability distribution so $\sum_i p_i = 1$ always. By equation (6.93a), $\ln p_i = -E_i/(k_\mathrm{B}T) - \ln Z$, so again using $\sum_i p_i = 1$, equation (6.104) can be rewritten

$$T\mathrm{d}S = \sum_i E_i \mathrm{d}p_i. \tag{6.105}$$

If we heat our system up at constant volume, the E_i stay the same but the p_i change because they depend on T. In these circumstances the increase in internal energy, $\sum_i E_i \mathrm{d}p_i$, is the heat absorbed by the system. Consequently, equation (6.105) states that $T\mathrm{d}S$ is the heat absorbed when the system is heated with no work done. This statement coincides with the definition of entropy in classical thermodynamics.

Substituting equation (6.105) into equation (6.103) yields

$$\mathrm{d}U = T\mathrm{d}S - P\mathrm{d}\mathcal{V}, \tag{6.106a}$$

where

$$P \equiv -\sum_i p_i \frac{\partial E_i}{\partial \mathcal{V}}. \tag{6.106b}$$

If we isolate our system from heat sources and then slowly change its volume, the adiabatic principle (§11.1) tells us that the system will stay in whatever stationary state it started in. That is, the p_i will be constant while the volume of the thermally isolated system is slowly changed. In classical thermodynamics this is an 'adiabatic' change. From equation (6.104) we see that the entropy S is constant during an adiabatic change, just as classical thermodynamics teaches.

Since $\mathrm{d}S = 0$ in an adiabatic change, the change in U as $\mathcal{V}$ is varied must be the mechanical work done on the system, $-P\mathrm{d}\mathcal{V}$, where P is the pressure the system exerts. This argument establishes that the quantity P defined by (6.106b) is the pressure.

Differentiating equation (6.100) for the Helmholtz free energy and using equation (6.106a) to eliminate $\mathrm{d}U$, we find that

$$\mathrm{d}F = -S\mathrm{d}T - P\mathrm{d}\mathcal{V}. \tag{6.107}$$

From this it immediately follows that

$$S = -\left(\frac{\partial F}{\partial T}\right)_\mathcal{V} \quad ; \quad P = -\left(\frac{\partial F}{\partial \mathcal{V}}\right)_T. \tag{6.108}$$

The first of these equations was obtained above but the second one is new.

Equation (6.106a) is the central equation of thermodynamics since it embodies both the first and second laws of thermodynamics. This result establishes that classical thermodynamics is a consequence of applying quantum mechanics to systems of which we know very little. Remarkably, physicists working in the first half of the 19^th century discovered thermodynamics long before quantum mechanics was thought of, using extremely subtle arguments concerning heat engines. Quantum mechanics makes these arguments redundant. Notwithstanding this redundancy, they continue to feature in undergraduate syllabuses the world over because they are beautiful. But then so are copperplate writing and slide rules, which have rightly disappeared from schools.

A possible explanation for the survival of thermodynamics as an independent discipline is as follows. Equations (6.99), (6.100) and (6.108) establish that any thermodynamic quantity can be obtained from the dependence of the partition function on T and V. Unfortunately, this dependence can be calculated for only a very few Hamiltonians. In almost all practical cases we cannot proceed by evaluating Z. However, once we know that Z and therefore F and S *exist*, we can determine their functional forms from experimental data. For example, by measuring the heat released on cooling our system at constant volume to absolute zero, we can determine its entropy $S = \int \mathrm{d}Q/T$. Similarly, we can measure the system's pressure as a function of T and $\mathcal{V}$. Then by integrating equation (6.107) we can obtain $F(T, \mathcal{V})$ and thus infer $Z(T, \mathcal{V})$. In none of these operations is the involvement of quantum mechanics apparent, so engineers and chemists, who make extensive use of thermodynamics, are generally unaware that it is a consequence of quantum mechanics. Quantum mechanics provides us with relations between thermodynamics quantities but does not enable us to evaluate the quantities themselves. Evaluation must still be done with 19th century technology.

Although thermodynamics systems are inherently macroscopic, quantum mechanics plays a central role in determining their thermodynamic quantities because it defines the stationary states we have to sum over in (6.93b) to form the partition function. Before quantum mechanics was born, the thermodynamic properties of an **ideal gas** – one composed of molecules that occupy negligible volume and interact only at very short range – were obtained by summing over the phase-space locations of each molecule independently. In this procedure there are six distinct states of a three-molecule gas in which there are molecules at the phase-space locations $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$: in one state molecule 1 is at $\mathbf{x}_1$, molecule 2 is at $\mathbf{x}_2$ and molecule 3 is at $\mathbf{x}_3$, and a distinct state is obtained by swapping the locations of molecule 1 and molecule 2, and so forth. Quantum mechanics teaches that the state of the gas is completely specified by listing the three occupied states, $|1\rangle$, $|2\rangle$ and $|3\rangle$ for it is meaningless to say which molecule is in which state. The classical way of counting states leads to absurd results even for gases at room temperature (Problem 6.17). At low-temperatures another aspect of classical physics leads to erroneous results: the low-lying energy levels of a gas are distributed discretely rather than continuously in E, with the result that specific heats always vanish in the limit $T \to 0$ (**Nernst's theorem**; Problem 6.18), contrary to the prediction of classical physics.

An important lesson to be learnt from the failure of classical physics to predict the properties of an ideal gas is the importance in quantum mechanics of thinking wholistically: we have to sum over the quantum states of the whole cylinder of gas, not over the states of individual molecules. This is analogous to the importance for understanding EPR phenomena of considering the quantum system formed by the entangled particles taken together. In quantum mechanics the whole is generally very much more than the sum of its parts because there are non-trivial correlations between the parts.[11]

---

[11] The origin of these correlations is the subject of §10.1.

6.5 Measurement

In §1.4 we asserted that the state of a system 'collapses' into one of the eigenstates $|q_j\rangle$ of the operator Q the instant we measure the observable Q. Consequently, the result of measuring Q is to leave the system in the well-defined quantum state $|q_j\rangle$. It's time to examine this **collapse hypothesis** critically.

Superficially the collapse hypothesis is merely an assertion that measurements are reproducible in the sense that if we measure something twice in quick succession, we will obtain the same result: in §2.1 $|q_j\rangle$ was *defined* to be the state in which measurement of Q was certain to yield the value q_j, so if the measurement of Q is to be reproducible, the system *has* to be in the state $|q_j\rangle$ immediately after the measurement. However, our system's quantum state $|\psi\rangle$ is supposed to describe the system's real, physical state, not just our knowledge of it. So something physical must have happened to make $|\psi\rangle$ shift from the value it had before the measurement to the state $|q_j\rangle \neq |\psi\rangle$ that it had just after the measurement was completed. Notice that the evolution from $|\psi\rangle$ to $|q_j\rangle$ has not been derived from the TDSE, which we have stated to be the equation that governs the time-evolution of $|\psi\rangle$. So this **Copenhagen interpretation** of quantum mechanics implies that every measurement leads to a momentary suspension of the equations of motion, so the system can be steered, by forces unspecified, into a randomly chosen state! This *not* serious physics. We need to consider more realistically what is involved in making a measurement.

To make a measurement, we need an instrument. The instrument is itself a dynamical system, and its dynamics is governed by quantum mechanics. We make a measurement by putting the instrument 'into contact' with our system – that is, we ensure that the instrument and the system are dynamically coupled by a non-negligible Hamiltonian. Once in contact, the instrument and our system together form a composite system, and like all dynamically coupled subsystems, they soon become entangled. That is, the state of the instrument becomes correlated with that of the system. This is promising because a measuring instrument should show a reading that is controlled by the system being measured. However, there is a big problem: the evolution of the state of the composite system that the TDSE provides is causal and leaves nothing to chance. Hence the entanglement of the states of the system and the instrument is causal, and if it is all that is involved in measurement, the result of the measurement will not be a matter of chance, contrary to the fundamental principles of quantum mechanics.

Our study of reduced density operators for subsystems suggests a resolution of this problem: if after it has become entangled with our system we treat the instrument as an independent system, it will be in an impure state and we can calculate classical probabilities that it is in any one of a complete set of basis states. To illustrate this idea, suppose both the system A and the instrument B are qubits, so they have just two basis states. Then the state of the system-plus-instrument might be given by equation (6.79), and taking the partial trace over the states of A, we obtain, by analogy with equation (6.80), that the instrument's reduced density matrix would be

$$\rho_{\mathrm{B}} = \tfrac{1}{2}\left(|\mathrm{B};0\rangle\langle\mathrm{B};0| + |\mathrm{B};1\rangle\langle\mathrm{B};1|\right). \tag{6.109}$$

In this case the (classical) probabilities that the instrument is in either of its basis states is $\frac{1}{2}$. Given our knowledge of the dynamics and thus equation (6.79), if we find the instrument to be in the state $|\mathrm{B};0\rangle$ we conclude that the system is in the state $|\mathrm{A};0\rangle$ and similarly if we find the instrument to be in the state $|\mathrm{B};1\rangle$. Unfortunately several problems stand in the way of extending this idea into a consistent general theory of quantum measurement. A major issue is that in general the instrument's reduced density operator will not be diagonal and the off-diagonal components of a density matrix are not classical probabilities.

Remarkably, more than 80 years after quantum mechanics became firmly established as a largely valid theory, there is still no satisfactory quantum theory of measurement.

Problems

6.1 A system AB consists of two non-interacting parts A and B. The dynamical state of A is described by $|a\rangle$, and that of B by $|b\rangle$, so $|a\rangle$ satisfies the TDSE for A and similarly for $|b\rangle$. What is the ket describing the dynamical state of AB? In terms of the Hamiltonians $H_{\rm A}$ and $H_{\rm B}$ of the subsystems, write down the TDSE for the evolution of this ket and show that it is automatically satisfied. Do $H_{\rm A}$ and $H_{\rm B}$ commute? How is the TDSE changed when the subsystems are coupled by a small dynamical interaction $H_{\rm int}$? If A and B are harmonic oscillators, write down $H_{\rm A}$, $H_{\rm B}$. The oscillating particles are connected by a weak spring. Write down the appropriate form of the interaction Hamiltonian $H_{\rm int}$. Does $H_{\rm A}$ commute with $H_{\rm int}$? Explain the physical significance of your answer.

6.2 Given that the state $|{\rm AB}\rangle$ of a compound system can be written as a product $|{\rm A}\rangle|{\rm B}\rangle$ of states of the individual systems, show that when $|{\rm AB}\rangle$ is written as $\sum_{ij} c_{ij}|{\rm A};i\rangle|{\rm B};j\rangle$ in terms of arbitrary basis vectors for the subsystems, every column of the matrix c_{ij} is a multiple of the leftmost column.

6.3 Consider a system of two particles of mass m that each move in one dimension along a given rod. Let $|1;x\rangle$ be the state of the first particle when it's at x and $|2;y\rangle$ be the state of the second particle when it's at y. A complete set of states of the pair of particles is $\{|xy\rangle\} = \{|1;x\rangle|2;y\rangle\}$. Write down the Hamiltonian of this system given that the particles attract one another with a force that's equal to C times their separation.

Suppose the particles experience an additional potential

$$V(x,y) = \tfrac{1}{2}C(x+y)^2. \tag{6.110}$$

Show that the dynamics of the two particles is now identical with the dynamics of a single particle that moves in two dimensions in a particular potential $\Phi(x,y)$, and give the form of Φ.

6.4 In §6.1.4 we derived Bell's inequality by considering measurements by Alice and Bob on an entangled electron-positron pair. Bob measures the component of spin along an axis that is inclined by angle θ to that used by Alice. Given the expression

$$|-,\mathbf{b}\rangle = \cos(\theta/2)\,{\rm e}^{{\rm i}\phi/2}|-\rangle - \sin(\theta/2)\,{\rm e}^{-{\rm i}\phi/2}|+\rangle, \tag{6.111}$$

for the state of a spin-half particle in which it has spin $+\frac{1}{2}$ along the direction $\mathbf{b}$ with polar angles (θ,ϕ), with $|\pm\rangle$ the states in which there is spin $\pm\frac{1}{2}$ along the z-axis, calculate the amplitude $A_{\rm B}(-|{\rm A}+)$ that Bob finds the positron's spin to be $-\frac{1}{2}$ given that Alice has found $+\frac{1}{2}$ for the electron's spin. Hence show that $P_{\rm B}(-|{\rm A}+) = \cos^2(\theta/2)$.

6.5 Show that when the Hadamard operator $U_{\rm H}$ is applied to every qubit of an n-qubit register that is initially in a member $|m\rangle$ of the computational basis, the resulting state is

$$|\psi\rangle = \frac{1}{2^{n/2}}\sum_{x=0}^{2^n-1} a_x|x\rangle, \tag{6.112}$$

where $a_x = 1$ for all x if $m = 0$, but exactly half the $a_x = 1$ and the other half the $a_x = -1$ for any other choice of m. Hence show that

$$\frac{1}{2^{n/2}}U_{\rm H}\sum_x a_x|x\rangle = \begin{cases} |0\rangle & \text{if all } a_x = 1 \\ |m\rangle \neq |0\rangle & \text{if half the } a_x = 1 \text{ and the other } a_x = -1. \end{cases} \tag{6.113}$$

6.6 Show that the trace of every Hermitian operator is real.

6.7 Let ρ be the density operator of a two-state system. Explain why ρ can be assumed to have the matrix representation

$$\rho = \begin{pmatrix} a & c \\ c^* & b \end{pmatrix}, \tag{6.114}$$

where a and b are real numbers. Let E_0 and $E_1 > E_0$ be the eigenenergies of this system and $|0\rangle$ and $|1\rangle$ the corresponding stationary states. Show from the equation of motion of ρ that in the energy representation a and b are time-independent while $c(t) = c(0)\mathrm{e}^{\mathrm{i}\omega t}$ with $\omega = (E_1 - E_0)/\hbar$.

Determine the values of a, b and $c(t)$ for the case that initially the system is in the state $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Given that the parities of $|0\rangle$ and $|1\rangle$ are even and odd respectively, find the time evolution of the expectation value $\overline{x}$ in terms of the matrix element $\langle 0|x|1\rangle$. Interpret your result physically.

6.8 Show that when the density operator takes the form $\rho = |\psi\rangle\langle\psi|$, the expression $\overline{Q} = \mathrm{Tr}\, Q\rho$ for the expectation value of an observable can be reduced to $\langle\psi|Q|\psi\rangle$. Explain the physical significance of this result. For the given form of the density operator, show that the equation of motion of ρ yields

$$|\phi\rangle\langle\psi| = |\psi\rangle\langle\phi| \quad \text{where} \quad |\phi\rangle \equiv \mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial t} - H|\psi\rangle. \tag{6.115}$$

Show from this equation that $|\phi\rangle = a|\psi\rangle$, where a is real. Hence determine the time evolution of $|\psi\rangle$ given the at $t = 0$ $|\psi\rangle = |E\rangle$ is an eigenket of H. Explain why ρ does not depend on the phase of $|\psi\rangle$ and relate this fact to the presence of a in your solution for $|\psi, t\rangle$.

6.9 The density operator is defined to be $\rho = \sum_\alpha p_\alpha |\alpha\rangle\langle\alpha|$, where p_α is the probability that the system is in the state α. Given an arbitrary basis $\{|i\rangle\}$ and the expansions $|\alpha\rangle = \sum_i a_{\alpha i}|i\rangle$, calculate the matrix elements $\rho_{ij} = \langle i|\rho|j\rangle$ of ρ. Show that the diagonal elements ρ_{ii} are non-negative real numbers and interpret them as probabilities.

6.10 Consider the density operator $\rho = \sum_{ij} \rho_{ij}|i\rangle\langle j|$ of a system that is in a pure state. Show that every row of the matrix ρ_{ij} is a multiple of the first row and every column is a multiple of the first column. Given that these relations between the rows and columns of a density matrix hold, show that the system is in a pure state. Hint: exploit the real, non-negativity of ρ_{11} established in Problem 6.9 and the Hermiticity of ρ.

6.11 Consider the rate of change of the expectation of the observable Q when the system is in an impure state. This is

$$\frac{\mathrm{d}\overline{Q}}{\mathrm{d}t} = \sum_n p_n \frac{\mathrm{d}}{\mathrm{d}t}\langle n|Q|n\rangle, \tag{6.116}$$

where p_n is the probability that the system is in the state $|n\rangle$. By using Ehrenfest's theorem to evaluate the derivative on the right of (6.116), derive the equation of motion $\mathrm{i}\hbar\mathrm{d}\overline{Q}/\mathrm{d}t = \mathrm{Tr}(\rho[Q, H])$.

6.12 Find the probability distribution $(p_1, \ldots, p_n)$ for n possible outcomes that maximises the Shannon entropy. Hint: use a Lagrange multiplier.

6.13 Use Lagrange multipliers λ and β to extremise the Shannon entropy of the probability distribution $\{p_i\}$ subject to the constraints (i) $\sum_i p_i = 1$ and (ii) $\sum_i p_i E_i = U$. Explain the physical significance of your result.

6.14 A composite system is formed from uncorrelated subsystem A and subsystem B, both in impure states. The numbers $\{p_{\mathrm{A}i}\}$ are the probabilities of the members of the complete set of states $\{|\mathrm{A}; i\rangle\}$ for subsystem A, while the numbers $\{p_{\mathrm{B}i}\}$ are the probabilities of the complete set of states $\{|\mathrm{B}; i\rangle\}$ for subsystem B. Show that the Shannon entropy of the composite system is the sum of the Shannon entropies of its subsystems. What is the relevance of this result for thermodynamics?

6.15 The $|0\rangle$ state of a qubit has energy 0, while the $|1\rangle$ state has energy ϵ. Show that when the qubit is in thermodynamic equilibrium at temperature $T = 1/(k_\mathrm{B}\beta)$ the internal energy of the qubit is

$$U = \frac{\epsilon}{\mathrm{e}^{\beta\epsilon} + 1}. \tag{6.117}$$

Show that when $\beta\epsilon \ll 1$, $U \simeq \frac{1}{2}\epsilon$, while for $\beta\epsilon \gg 1$, $U \simeq \epsilon\mathrm{e}^{-\beta\epsilon}$. Interpret these results physically and sketch the specific heat $C = \partial U/\partial T$ as a function of T.

6.16 Show that the partition function of a harmonic oscillator of natural frequency ω is

$$Z_\mathrm{ho} = \frac{\mathrm{e}^{-\beta\hbar\omega/2}}{1 - \mathrm{e}^{-\beta\hbar\omega}}. \tag{6.118}$$

Hence show that when the oscillator is at temperature $T = 1/(k_\mathrm{B}\beta)$ the oscillator's internal energy is

$$U_\mathrm{ho} = \hbar\omega\left(\frac{1}{2} + \frac{1}{\mathrm{e}^{\beta\hbar\omega} - 1}\right). \tag{6.119}$$

Interpret the factor $(\mathrm{e}^{\beta\hbar\omega} - 1)^{-1}$ physically. Show that the specific heat $C = \partial U/\partial T$ is

$$C = k_\mathrm{B}\frac{\mathrm{e}^{\beta\hbar\omega}}{(\mathrm{e}^{\beta\hbar\omega} - 1)^2}(\beta\hbar\omega)^2. \tag{6.120}$$

Show that $\lim_{\mathrm{T}\to 0} C = 0$ and obtain a simple expression for C when $k_\mathrm{B}T \gg \hbar\omega$.

6.17 A classical ideal monatomic gas has internal energy $U = \frac{3}{2}Nk_\mathrm{B}T$ and pressure $P = Nk_\mathrm{B}T/\mathcal{V}$, where N is the number of molecules and $\mathcal{V}$ is the volume they occupy. From these relations, and assuming that the entropy vanishes at zero temperature and volume, show that in general the entropy is

$$S(T, \mathcal{V}) = Nk_\mathrm{B}(\tfrac{3}{2}\ln T + \ln \mathcal{V}). \tag{6.121}$$

A removable wall divides a cylinder into equal parts of volume $\mathcal{V}$. Initially the wall is in place and each half contains N molecules of ideal monatomic gas at temperature T. The wall is removed. Show that equation (6.121) implies that the entropy of the entire body of fluid increases by $2\ln 2\, Nk_\mathrm{B}$. Can this result be squared with the principle that $\mathrm{d}S = \mathrm{d}Q/T$, where $\mathrm{d}Q$ is the heat absorbed when the change is made reversibly? What conclusion do you draw from this thought experiment?

6.18 Consider a 'gas' formed by M non-interacting, monatomic molecules of mass m that move in a one-dimensional potential well $V = 0$ for $|x| < a$ and ∞ otherwise. Assume that at sufficiently low temperatures all molecules are either in the ground or first-excited states. Show that in this approximation the partition function is given by

$$\ln Z = -M\beta E_0 + \mathrm{e}^{-3\beta E_0} - \mathrm{e}^{-3(M+1)\beta E_0} \quad \text{where} \quad E_0 \equiv \frac{\pi^2\hbar^2}{8ma^2}. \tag{6.122}$$

Show that for M large the internal energy, pressure and specific heat of this gas are given by

$$U = E_0(M + 3\mathrm{e}^{-3\beta E_0}) \ ; \ P = \frac{2E_0}{a}\left(M + 3\mathrm{e}^{-3\beta E_0}\right) \ ; \ C_V = \frac{9E_0^2}{k_\mathrm{B}T^2}\mathrm{e}^{-3\beta E_0}. \tag{6.123}$$

In what respects do these results for a quantum ideal gas differ from the properties of a classical ideal gas? Explain these differences physically.

7

Angular Momentum

In Chapter 4 we introduced the angular-momentum operators J_i as the generators of rotations. We showed that they form a pseudo-vector, so $J^2 = \sum_i J_i^2$ is a scalar. By considering the effect of rotations on vectors and scalars, we showed that the the J_i commute with all scalar operators, including J^2, and found that commutator of J_i with a component of vector operator is given by equation (4.30). From this result we deduced that the J_i do not commute with one another, but satisfy $[J_i, J_j] = \mathrm{i}\sum_k \epsilon_{ijk} J_k$.

Although we have from the outset called the J_i 'angular-momentum operators', the only connection we have established between the J_i and angular momentum is tenuous and by no means justifies our terminology: we have simply shown that when the Hamiltonian is invariant under rotations about some axis $\hat{\boldsymbol{\alpha}}$, and the system starts in an eigenstate of the corresponding angular-momentum operator $\hat{\boldsymbol{\alpha}} \cdot \mathbf{J}$, it will subsequently remain in that eigenstate. Consequently, the corresponding eigenvalue is then a conserved quantity. In classical mechanics dynamical symmetry about some axis implies that the component of angular momentum about that axis is conserved, so it is plausible that the conserved eigenvalue is a measure of angular momentum. This suggestion will be substantiated in this chapter. Another important task for the chapter is to explain how the orientation of a system is encoded in the amplitudes for it to be found in different eigenstates of appropriate angular-momentum operators. We start by using the angular-momentum commutation relations to determine the spectrum of the J_i.

7.1 Eigenvalues of J_z and J^2

Since no two components of $\mathbf{J}$ commute, we cannot find a complete set of simultaneous eigenkets of two components of $\mathbf{J}$. We can, however, find a complete set of mutual eigenkets of J^2 and *one* component of $\mathbf{J}$ because $[J^2, J_i] = 0$. Without loss of generality we can orient our coordinates so that the chosen component of $\mathbf{J}$ is J_z. Let us label a ket which is simultaneously an eigenstate of J^2 and J_z as $|\beta, m\rangle$, where

$$J^2|\beta, m\rangle = \beta|\beta, m\rangle \quad ; \quad J_z|\beta, m\rangle = m|\beta, m\rangle. \tag{7.1}$$

We now define

$$J_\pm \equiv J_x \pm \mathrm{i}J_y. \tag{7.2}$$

These objects clearly commute with J^2, while their commutation relation with J_z are

$$[J_\pm, J_z] = [J_x, J_z] \pm \mathrm{i}[J_y, J_z] = -\mathrm{i}J_y \mp J_x = \mp J_\pm. \tag{7.3}$$

Since $J_\pm$ commutes with J^2, the ket $J_\pm|\beta, m\rangle$ is an eigenket of J^2 with eigenvalue β. Operating with J_z on this ket we find

$$J_z J_\pm|\beta, m\rangle = (J_\pm J_z + [J_z, J_\pm])\,|\beta, m\rangle = (m \pm 1) J_\pm|\beta, m\rangle. \tag{7.4}$$

Thus, $J_\pm|\beta, m\rangle$ is also a member of the complete set of states that are eigenstates of both J^2 and J_z, but its eigenvalue with respect to J_z differs from that of $|\beta, m\rangle$ by ± 1. Therefore we may write

$$J_\pm|\beta, m\rangle = \alpha_\pm|\beta, m \pm 1\rangle, \tag{7.5}$$

where $\alpha_\pm$ is a constant that we now evaluate. We do this by taking the length-squared of both sides of equation (7.5). Bearing in mind that $J_\pm^\dagger = J_\mp$, we find

$$\begin{aligned} |\alpha_\pm^2| &= \langle\beta, m|J_\mp J_\pm|\beta, m\rangle = \langle\beta, m|(J_x \mp \mathrm{i}J_y)(J_x \pm \mathrm{i}J_y)|\beta, m\rangle \\ &= \langle\beta, m|(J^2 - J_z^2 \mp J_z)|\beta, m\rangle = \{\beta - m(m \pm 1)\}, \end{aligned} \tag{7.6}$$

so

$$\alpha_\pm = \sqrt{\beta - m(m \pm 1)}. \tag{7.7}$$

The J_i are Hermitian operators, so $\langle\psi|J_i^2|\psi\rangle = \big|J_i|\psi\rangle\big|^2 \geq 0$. Hence

$$\beta = \langle\beta, m|J^2|\beta, m\rangle = \langle\beta, m|(J_x^2 + J_y^2 + J_z^2)|\beta, m\rangle \geq m^2. \tag{7.8}$$

So notwithstanding equation (7.5), it cannot be possible to create states with ever larger eigenvalues of J_z by repeated application of J_+. All that can stop us doing this is the vanishing of α_+ when we reach some maximum eigenvalue $m_{\max}$ that from equation (7.7) satisfies

$$\beta - m_{\max}(m_{\max} + 1) = 0. \tag{7.9}$$

Similarly, α_- must vanish for a smallest value of m that satisfies

$$\beta - m_{\min}(m_{\min} - 1) = 0. \tag{7.10}$$

Eliminating β between (7.9) and (7.10) we obtain a relation between $m_{\max}$ and $m_{\min}$ that we can treat as a quadratic equation for $m_{\min}$. Solving this equation we find that

$$m_{\min} = \tfrac{1}{2}\{1 \pm (2m_{\max} + 1)\}. \tag{7.11}$$

The plus sign yields a value of $m_{\min}$ that is incompatible with our requirement that $m_{\min} \leq m_{\max}$, so we must have $m_{\min} = -m_{\max}$. To simplify the notation, we define $j \equiv m_{\max}$, so that equation (7.9) becomes $\beta = j(j+1)$ and $-j \leq m \leq j$. Finally, we note that since an integer number of applications of J_- will take us from $|\beta, j\rangle$ to $|\beta, -j\rangle$, $2j$ must be an integer – see Figure 7.1. In summary, the eigenvalues of J^2 are $j(j+1)$ with $2j = 0, 1, 2, \ldots$ and for each value of j the eigenvalues m of J_z are $(j, j-1, \ldots, -j)$.

At this point we simplify the labelling of kets by defining $|j, m\rangle$ to be what has hitherto been denoted $|\beta, m\rangle$ with $\beta = j(j+1)$ – we clear a great deal of clutter from the page by replacing $|j(j+1), m\rangle$ with $|j, m\rangle$. The kets' eigenvalues with respect to J^2 are of course unaffected by this relabelling.

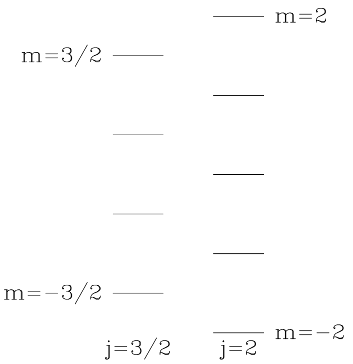


Figure 7.1 Going from $m_{\rm min}$ to $m_{\rm max}$ in an integer number of steps in the cases $j = \frac{3}{2}, 2$.

Had we known at the outset that the eigenvalues of J^2 would be of the form $j(j+1)$, we would have used the new notation all along.

In summary, we can find simultaneous eigenstates of J^2 and one of the J_i, conventionally taken to be J_z. The eigenvalues of J^2 are $j(j+1)$ with $2j = 0, 1, \ldots$, and for any given j the eigenvalues m of J_z then run from $+j$ to $-j$ in integer steps:

$$j \geq m \geq -j. \tag{7.12}$$

In order to move from the state $|j,m\rangle$ to the adjacent state $|j,m\pm1\rangle$ we use the raising or lowering operators $J_\pm$ which act as

$$J_\pm|j,m\rangle = \alpha_\pm(m)|j,m\pm1\rangle = \sqrt{j(j+1)-m(m\pm1)}|j,m\pm1\rangle. \tag{7.13}$$

These operators only change the J_z eigenvalue, so they just realign a given amount of total angular momentum, placing more (J_+) or less (J_-) along the z-axis. So far, we have not discovered how to alter the J^2 eigenvalue $j(j+1)$.

It is sometimes helpful to rewrite the constants $\alpha_\pm(m)$ of equation (7.13) in the form

$$\begin{aligned}\alpha_+(m) &= \sqrt{(j-m)(j+m+1)}\\ \alpha_-(m) &= \sqrt{(j+m)(j-m+1)}.\end{aligned} \tag{7.14}$$

These equations make it clear that the proportionality constants for different m satisfy

$$\begin{aligned}\alpha_+(m) &= \alpha_+(-m-1) \qquad & \alpha_+(m-1) &= \alpha_-(m)\\ \alpha_-(m) &= \alpha_-(-m+1) \qquad & \alpha_-(m) &= \alpha_+(-m).\end{aligned} \tag{7.15}$$

For example, when J_- lowers the highest state $|j,j\rangle$, we obtain the same proportionality constant as when J_+ raises the lowest state $|j,-j\rangle$; consequently, we only need to work out half the constants directly, because we can then infer the others.

In §4.1.2 we discovered that when the system is rotated through an angle α around the z axis, its ket $|\psi\rangle$ transforms to $|\psi'\rangle = U(\alpha)|\psi\rangle$, where the unitary operator $U(\alpha) = \exp(-{\rm i}\alpha J_z)$. If $|\psi\rangle = |j,m\rangle$ is an eigenket of J_z, $U(\alpha)$ simply changes its phase:

$$U(\alpha)|j,m\rangle = {\rm e}^{-{\rm i}\alpha J_z}|j,m\rangle = {\rm e}^{-{\rm i}\alpha m}|j,m\rangle. \tag{7.16}$$

Since $2j$ is an integer, j (and hence m) must be either an integer or a half integer. Using this information in equation (7.16), we see that, after a rotation through 2π around the z-axis, we have either

$$|j,m\rangle \to |j,m\rangle \quad \text{for } m \text{ even} \tag{7.17a}$$

or

$$|j,m\rangle \to -|j,m\rangle \quad \text{for } m \text{ odd.} \tag{7.17b}$$

Equation (7.17a) is as expected; under a 2π rotation, the system returns to its original state. However, equation (7.17b) says that a system with half integer angular momentum does not return to its original state after a 2π rotation – the initial and final states are minus one another! This difference of behaviour between systems with integer and half-integer angular momentum is of fundamental importance, and determines many other characteristics of these systems. A result of quantum field theory is that 'spin-half' fields never attain macroscopic values: the quantum uncertainty in the value of a spin-half field is always on the same order as the value of the field itself. Integer-spin fields, by contrast, can attain macroscopic values: values that are vastly greater than their quantum uncertainties. Consequently, classical physics – physics in the absence of quantum uncertainty – involves integer-spin fields (the electromagnetic and gravitational fields are examples) but no spin-half field. Our intuition about what happens when a system is rotated has grown out of our experience of classical physics, so we consider that things return to their original state after rotation by 2π. If we had hands-on experience of spin-half objects, we would recognise that this is not generally true.

7.1.1 Rotation spectra of diatomic molecules

Knowledge of the spectrum of the angular momentum operators enables us to understand an important part of the dynamics of a diatomic molecule such as carbon monoxide. For some purposes a CO molecule can be considered to consist of two point masses, the nuclei of the oxygen and carbon atoms, joined by a 'light rod' provided by the electrons. In this model the molecule's moment of inertia around the axis that joins the nuclei is negligible, while the same moment of inertia I applies to any perpendicular axis.

In classical mechanics the rotational energy of a rigid body is

$$E = \tfrac{1}{2}\left(\frac{\mathcal{J}_x^2}{I_x} + \frac{\mathcal{J}_y^2}{I_y} + \frac{\mathcal{J}_z^2}{I_z}\right), \tag{7.18}$$

where the I_i are the moments of inertia about the body's three principal axes and $\boldsymbol{\mathcal{J}}$ is the body's angular-momentum vector. We conjecture that the equivalent formula links the Hamiltonian and the angular momentum operators in quantum mechanics:

$$H = \frac{\hbar^2}{2}\left(\frac{J_x^2}{I_x} + \frac{J_y^2}{I_y} + \frac{J_z^2}{I_z}\right). \tag{7.19}$$

The best justification for adopting this formula is that it leads us to results that are confirmed by experiments.

In the case of an axisymmetric body, we orient our body such that the symmetry axis is parallel to the z axis. Then $I \equiv I_x = I_y$ and the Hamiltonian can be written

$$H = \frac{\hbar^2}{2}\left\{\frac{J^2}{I} + J_z^2\left(\frac{1}{I_z} - \frac{1}{I}\right)\right\}. \tag{7.20}$$

From this formula and our knowledge of the eigenvalues of J^2 and J_z, we can immediately write down the energies that form the spectrum of H:

$$E_{jm} = \frac{\hbar^2}{2}\left\{\frac{j(j+1)}{I} + m^2\left(\frac{1}{I_z} - \frac{1}{I}\right)\right\}, \tag{7.21}$$

where j is the total angular-momentum quantum number and $|m| < j$. In the case of a diatomic molecule such as CO, $I_z \ll I$ so the coefficient of m^2 is very much larger than the coefficient of $j(j+1)$ and states with $|m| > 0$ will

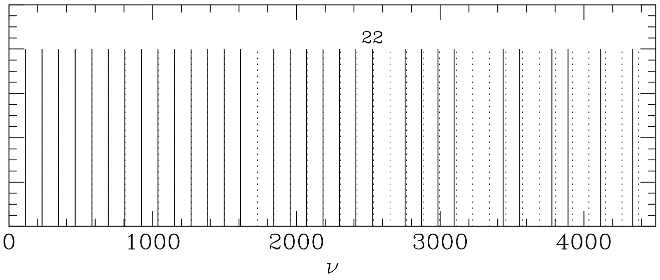


Figure 7.2 The rotation spectrum of CO. The full lines show the measured frequencies for transitions up to $j = 38 \rightarrow 37$, while the dotted lines show integer multiples of the lowest measured frequency. Up to the line for $j = 22 \rightarrow 21$ the dotted lines are obscured by the full lines except at one frequency for which measurements are not available. For $j \geq 22$ the separation between the dotted and full lines increases steadily as a consequence of the centrifugal stretching of the bond between the molecule's atoms. Measurements are lacking for several of the higher-frequency lines.

occur only far above the ground state. Consequently, the states of interest have energies of the form

$$E_j = j(j+1)\frac{\hbar^2}{2I}. \tag{7.22}$$

For reasons that will become clearer in §7.2.2, only integer values of j are allowed.

CO is a significantly dipolar molecule. The carbon atom has a smaller share of the binding electrons than the oxygen atom, with the result that it is positively charged and the oxygen atom is negatively charged. A rotating electric dipole would be expected to emit electromagnetic radiation. Because we are in the quantum regime, the radiation emerges as photons which, as we shall see, can add or carry away only one unit $\hbar$ of angular momentum. It follows that the energies of the photons that can be emitted or absorbed by a rotating dipolar molecule are

$$E_p = \pm\left(E_j - E_{j-1}\right) = \pm j\frac{\hbar^2}{I}. \tag{7.23}$$

Using the relation $E = h\nu$ between the energy of a photon and the frequency ν of its radiation, the frequencies in the rotation spectrum of the molecule are

$$\nu_j = j\frac{\hbar}{2\pi I}. \tag{7.24}$$

In the case of ^{12}CO, the coefficient of j evaluates to 113.1724 GHz and spectral lines occur at multiples of this frequency (Figure 7.2).

In the classical limit of large j, $\mathcal{J} = j\hbar$ is the molecule's angular momentum, and this is related to the angular frequency ω at which the molecule rotates by $\mathcal{J} = I\omega$. When in equation (7.24) we replace $j\hbar$ by $I\omega$, we discover that the frequency of the emitted radiation ν is simply the frequency $\omega/2\pi$ at which the molecule rotates around its axis. This conclusion makes perfect sense physically. Now, because of the form of the Hamiltonian, the energy eigenstates are also the eigenstates of J_z and J^2. Therefore in any energy eigenstate, $\left\langle J^2\right\rangle = j(j+1)$ and for low-lying states with $m = 0$ and $j \sim \mathcal{O}(1)$, $j(j+1)$ is significantly larger than j^2. Therefore ν_j in (7.24) is smaller than the frequency at which the molecule rotates when it is in the upper state of the transition. On the other hand, ν_j is larger than the rotation frequency $\sqrt{(j-1)j}\frac{\hbar}{2\pi I}$ of the lower state. Hence the frequency at which radiation emerges lies between the rotation frequencies of the upper and lower states. Again this makes sense physically. As we approach the classical regime, j becomes large so $j(j+1) \simeq j^2 \simeq (j-1)j$ and the rotation frequencies of the upper and lower states converge, from above and below, on the frequency of the emitted radiation.

Measurements of radiation from 115 GHz and the first few multiples of this frequency provide one of the two most important probes of interstellar gas.[1] In denser, cooler regions, hydrogen atoms combine to form H_2 molecules, which are bisymmetric and do not have an electric dipole moment when they are simply rotating. Consequently, these molecules, which together with similarly uncommunicative helium atoms make up the great majority of the mass of cold interstellar gas, lack readily observable spectral lines. Hence astronomers are obliged to study the cold interstellar medium through the rotation spectrum of the few parts in 10^6 of CO that it contains.

Important information can be gleaned from the relative intensities of lines associated with different values of j in equation (7.24). The rate at which molecules emit radiation and thus the intensity of the line[2] is proportional to the number n_j of molecules in the upper state. As we shall deduce in §7.5.3, all states have equal a priori probability, so n_j is proportional to the number of states that have the given energy – the **degeneracy** or **statistical weight** g of the energy level. From §7.1 we know that $g = 2j + 1$ because this is the number of possible orientations of the angular momentum for quantum number j.

In §6.4 we saw that when a gas is in thermal equilibrium at temperature T, the probability p_j that a given molecule is in a state of energy E_j is proportional to the Boltzmann factor $\exp(-E_j/k_\mathrm{B}T)$, where k_B is the Boltzmann constant (eq. 6.93a). Combining this proportionality with the dependence on the degeneracy $2j + 1$ just discussed leads us to expect that the intensity of the line at frequency ν_j will be

$$\mathcal{I}_j \propto (2j+1)\exp(-E_j/k_\mathrm{B}T) \qquad (j>0). \tag{7.25}$$

For $E_1 < k_\mathrm{B}T$, $\mathcal{I}_j$ increases at small j before declining as the Boltzmann factor begins to overwhelm the degeneracy factor. Fitting this formula, which has only one free parameter (T), to observed line intensities enables one both to measure the temperature of the gas, and to check the correctness of the degeneracy factor.

Figure 7.2 shows that for large values of the quantum number j, the spacing between lines in the spectrum diminishes in apparent violation of the prediction of equation (7.24). Lines with large j are generated by molecules that are spinning very rapidly. The bond between the nuclei is stretched like a spring by the centripetal acceleration of the nuclei. Stretching of the bond increases the moment of inertia I, and from equation (7.24) this decreases the frequency of the spectral lines (Problem 7.2).

7.2 Orbital angular momentum

Let $\mathbf{x}$ and $\mathbf{p}$ be the position and momentum operators of the system. Then, inspired by classical mechanics, we define the dimensionless **orbital angular momentum operators** by[3]

$$\mathbf{L} \equiv \frac{1}{\hbar}\mathbf{x}\times\mathbf{p}, \quad \text{that is} \quad L_i \equiv \frac{1}{\hbar}\sum_{jk}\epsilon_{ijk}x_jp_k. \tag{7.26}$$

From the rules of Table 2.1 and the Hermitian nature of $\mathbf{x}$ and $\mathbf{p}$, the Hermitian adjoint of L_i is

$$L_i^\dagger = \frac{1}{\hbar}\sum_{jk}\epsilon_{ijk}p_k^\dagger x_j^\dagger = \frac{1}{\hbar}\sum_{jk}\epsilon_{ijk}x_jp_k = L_i, \tag{7.27}$$

---

[1] The other key probe is the hyperfine line of atomic hydrogen that will be discussed in Chapter 8.

[2] We neglect the absorption of photons after emission, which can actually be an important process, especially for ^{12}CO.

[3] In many texts $\mathbf{L}$ is defined without the factor $\hbar^{-1}$. By making $\mathbf{L}$ dimensionless, this factor simplifies many subsequent formulae.

where we have used the fact that $[x_j, p_k] = 0$ for $j \neq k$. Thus the L_i are Hermitian and are likely to correspond to observables. We also define the **total orbital angular momentum operator** by

$$L^2 \equiv \mathbf{L} \cdot \mathbf{L} = L_x^2 + L_y^2 + L_z^2, \tag{7.28}$$

which is again Hermitian, and calculate a number of commutators. First, bearing in mind the canonical commutation relation (2.54), we have

$$\begin{aligned}[L_i, x_l] &= \frac{1}{\hbar}\sum_{jk} \epsilon_{ijk}[x_j p_k, x_l] = \frac{1}{\hbar}\sum_{jk} \epsilon_{ijk} x_j [p_k, x_l] = -\mathrm{i}\sum_j \epsilon_{ijl} x_j \\ &= \mathrm{i}\sum_j \epsilon_{ilj} x_j.\end{aligned} \tag{7.29}$$

Similarly

$$[L_i, p_l] = \frac{1}{\hbar}\sum_{jk} \epsilon_{ijk}[x_j p_k, p_l] = \frac{1}{\hbar}\sum_{jk} \epsilon_{ijk}[x_j, p_l] p_k = \mathrm{i}\sum_j \epsilon_{ilj} p_j. \tag{7.30}$$

Notice that these commutation relations differ from the corresponding ones for J_i [equations (4.29) and (4.31)] only by the substitution of L for J. From these relations we can show that L_i commutes with the scalars $\mathbf{x}^2$, $\mathbf{p}^2$ and $\mathbf{x} \cdot \mathbf{p}$. For example

$$[L_i, p^2] = \sum_j [L_i, p_j^2] = \mathrm{i}\sum_{jk} \epsilon_{ijk}(p_k p_j + p_j p_k) = 0, \tag{7.31}$$

where the last equality follows because the ϵ symbol is antisymmetric in jk while the bracket is symmetrical in these indices (see also Problem 7.1). We can now also calculate the commutator of one component of $\mathbf{L}$ with another. We have

$$[L_x, L_y] = \frac{1}{\hbar}[L_x, (zp_x - xp_z)] = \mathrm{i}(-yp_x + xp_y) = \mathrm{i}L_z. \tag{7.32}$$

Clearly each L_i commutes with itself, and the other non-zero commutators can be obtained from equation (7.32) by permuting indices. These commutators mirror the commutators (7.103) of the J_i.

$\mathbf{L}$ is a vector operator by virtue of the way it is constructed out of the vectors $\mathbf{x}$ and $\mathbf{p}$. It follows that L^2 is a scalar operator. Hence the way these operators commute with the total angular momentum operators J_i follows from the work of §4.2:

$$[J_i, L_j] = \mathrm{i}\sum_k \epsilon_{ijk} L_k \quad ; \quad [J_i, L^2] = 0. \tag{7.33}$$

Although p^2 and x^2 commute with L_i, the total angular momentum operator J^2 does not:

$$[J^2, L_i] = \sum_j [J_j^2, L_i] = \mathrm{i}\sum_{jk} \epsilon_{jik}(L_k J_j + J_j L_k). \tag{7.34}$$

The right side does not vanish because the final bracket is not symmetric in jk. The physical significance of $[J^2, L_i]$ being non-zero is that if our system is in a state of well-defined total angular momentum, in general there will be uncertainty in the amount of orbital angular momentum it has about *any* axis. We shall explore the consequences of this fact in §7.5.

Figure 7.3 J both swings the particle around the origin and rotates its spin (left), while L moves the particle, but leaves the direction of the spin invariant (right).

7.2.1 L as the generator of circular translations

In §4.1.1 we saw that when the system is displaced along a vector $\mathbf{a}$, its ket is transformed by the unitary operator $U(\mathbf{a}) = \mathrm{e}^{-\mathrm{i}\mathbf{a}\cdot\mathbf{p}/\hbar}$. We now imagine successively performing n translations through vectors $\{\mathbf{a}_1, \mathbf{a}_2 \ldots, \mathbf{a}_n\}$. Since each translation will cause $|\psi\rangle$ to be acted on by a unitary operator, the final state will be

$$\begin{aligned}U(\mathbf{a}_n)\ldots U(\mathbf{a}_2)U(\mathbf{a}_1)|\psi\rangle &= \left\{\prod_{i=1}^{n} \exp\left(\frac{\mathrm{i}}{\hbar}\mathbf{a}_i\cdot\mathbf{p}\right)\right\}|\psi\rangle \\ &= \exp\left\{\frac{\mathrm{i}}{\hbar}\left(\sum_{i=1}^{n}\mathbf{a}_i\right)\cdot\mathbf{p}\right\}|\psi\rangle,\end{aligned} \tag{7.35}$$

where the second equality follows because the components of $\mathbf{p}$ commute with one another. Since the exponent in the last line is proportional to the overall displacement vector $\mathbf{A} \equiv \sum_{i=1}^{n}\mathbf{a}_i$, the change in $|\psi\rangle$ is independent of the path that the system takes. In particular, if the path is closed, $\mathbf{A} = 0$ and $|\psi\rangle$ is unchanged.

Now consider the effect of moving the system in a circle centred on the origin and in the plane with unit normal $\mathbf{n}$. When we increment the rotation angle α by $\delta\alpha$, we move the system through

$$\delta\mathbf{a} = \delta\alpha\,\mathbf{n}\times\mathbf{x}. \tag{7.36}$$

The associated unitary operator is

$$\begin{aligned}U(\delta\mathbf{a}) &= \exp\left(-\frac{\mathrm{i}}{\hbar}\delta\alpha\,(\mathbf{n}\times\mathbf{x})\cdot\mathbf{p}\right) = \exp\left(-\frac{\mathrm{i}}{\hbar}\delta\alpha\,\mathbf{n}\cdot(\mathbf{x}\times\mathbf{p})\right) \\ &= \mathrm{e}^{-\mathrm{i}\delta\alpha\,\mathbf{n}\cdot\mathbf{L}}.\end{aligned} \tag{7.37}$$

The unitary operator corresponding to rotation through a finite angle α is a high power of this operator. Since the exponent contains only one operator, $\mathbf{n}\cdot\mathbf{L}$, which inevitably commutes with itself, the product of the exponentials is simply

$$U(\alpha) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{L}}, \tag{7.38}$$

where $\boldsymbol{\alpha} \equiv \alpha\mathbf{n}$.

The difference between the total and orbital angular momentum operators is now apparent. When we rotate the system on a turntable through an angle α, the system's ket is updated by $\mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J}}$. When we move the system around a circle *without modifying its orientation*, the ket is updated by $\mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{L}}$. The crucial insight is that the turntable both moves the system around a circle and reorientates it. The transformations of which $\mathbf{J}$ is the generator reflects both of these actions. The transformations of which $\mathbf{L}$ is the generator reflects only the translation.

7.2.2 Spectra of L^2 and L_z

We have shown that the L_i commute with one another in exactly the same way that the J_i do. In §7.1 we found the possible eigenvalues of J^2 and J_z from the commutation relations and nothing else. Hence we can without further ado conclude that the possible eigenvalues of L^2 and L_z are $l(l+1)$ and m, respectively, with $-l \leq m \leq l$, where l is a member of the set $(0, \frac{1}{2}, 1, \frac{3}{2}, \ldots)$.

In the last subsection we saw that $\mathbf{L}$ is the generator of translations on circles around the origin, and we demonstrated that when a complete rotation through 2π is made, the unitary operator that $\mathbf{L}$ generates is simply the identity. Consider the case in which we move the system right around the z axis when it is in the eigenstate $|l, m\rangle$ of L^2 and L_z. The unitary operator is then $\mathrm{e}^{-2\pi \mathrm{i} L_z}$ and the transformed ket is

$$|l, m\rangle = \mathrm{e}^{-2\pi \mathrm{i} L_z}|l, m\rangle = \mathrm{e}^{-2m\pi \mathrm{i}}|l, m\rangle. \tag{7.39}$$

Since the exponential on the right side is equal to unity only for integer m, we conclude that L_z, unlike J_z has only integer eigenvalues. Since for given l, m runs from $-l$ to l, it follows that l also takes only integer values. Thus the spectrum of L^2 is $l(l+1)$ with $l = 0, 1, 2, \ldots$, and for given l the possible values of L_z are the integers in the range $(-l, l)$.

7.2.3 Orbital angular momentum eigenfunctions

We already know the possible eigenvalues of the operators L^2 and L_z. Now we find the corresponding eigenfunctions.

In the position representation, the L_i become differential operators. For example

$$L_z = \frac{1}{\hbar}(xp_y - yp_x) = -\mathrm{i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) \tag{7.40}$$

Let (r, θ, ϕ) be standard spherical polar coordinates. Then the chain rule states that

$$\frac{\partial}{\partial \phi} = \frac{\partial x}{\partial \phi}\frac{\partial}{\partial x} + \frac{\partial y}{\partial \phi}\frac{\partial}{\partial y} + \frac{\partial z}{\partial \phi}\frac{\partial}{\partial z}. \tag{7.41}$$

Using $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$ and $z = r\cos\theta$ we find

$$\frac{\partial}{\partial \phi} = r\sin\theta\left(-\sin\phi\frac{\partial}{\partial x} + \cos\phi\frac{\partial}{\partial y}\right) = \left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) = \mathrm{i}L_z. \tag{7.42}$$

That is

$$L_z = -\mathrm{i}\frac{\partial}{\partial \phi}. \tag{7.43}$$

Let $|l, m\rangle$ be a simultaneous eigenket of L^2 and m for the eigenvalues $l(l+1)$ and m, respectively. Then $L_z|l, m\rangle = m|l, m\rangle$ and the wavefunction $\psi_{lm}(\mathbf{x}) \equiv \langle \mathbf{x}|l, m\rangle$ must satisfy the eigenvalue equation

$$-\mathrm{i}\frac{\partial \psi_{lm}}{\partial \phi} = m\psi_{lm}. \tag{7.44}$$

The solution of this equation is

$$\psi_{lm}(r, \theta, \phi) = K_{lm}(r, \theta)\mathrm{e}^{\mathrm{i}m\phi}, \tag{7.45}$$

where K_{lm} is an arbitrary function of r and θ. Since m is an integer, ψ_{lm} is a single-valued function of position.

In our determination of the spectra of J^2 and J_z in §7.1, important roles were played by the ladder operators $J_\pm = (J_x \pm \mathrm{i}J_y)$. If we define

$$L_\pm \equiv L_x \pm \mathrm{i}L_y, \tag{7.46}$$

then by analogy with equation (7.5) we will have that

$$L_\pm|l,m\rangle = \alpha_\pm|l,m\pm 1\rangle, \tag{7.47a}$$

where

$$\alpha_\pm(m) = \sqrt{l(l+1)-m(m\pm 1)}. \tag{7.47b}$$

It will be helpful to express $L_\pm$ in terms of partial derivatives with respect to spherical polar coordinates. We start by deriving a relation between partial derivatives that we will subsequently require. From the chain rule we have that

$$\frac{\partial}{\partial\theta} = r\cos\theta\Big(\cos\phi\frac{\partial}{\partial x}+\sin\phi\frac{\partial}{\partial y}\Big) - r\sin\theta\frac{\partial}{\partial z}. \tag{7.48a}$$

Multiplying the corresponding expression (7.42) for ϕ by $\cot\theta$ yields

$$\cot\theta\frac{\partial}{\partial\phi} = r\cos\theta\left(-\sin\phi\frac{\partial}{\partial x}+\cos\phi\frac{\partial}{\partial y}\right). \tag{7.48b}$$

Adding or subtracting i times (7.48b) to (7.48a) we obtain

$$\begin{aligned}\frac{\partial}{\partial\theta}\pm\mathrm{i}\cot\theta\frac{\partial}{\partial\phi} &= r\cos\theta\left\{(\cos\phi\mp\mathrm{i}\sin\phi)\frac{\partial}{\partial x}+(\sin\phi\pm\mathrm{i}\cos\phi)\frac{\partial}{\partial y}\right\} - r\sin\theta\frac{\partial}{\partial z}\\ &= r\cos\theta\mathrm{e}^{\mp\mathrm{i}\phi}\Big(\frac{\partial}{\partial x}\pm\mathrm{i}\frac{\partial}{\partial y}\Big) - r\sin\theta\frac{\partial}{\partial z}.\end{aligned} \tag{7.49}$$

Multiplying through by $\mathrm{e}^{\pm\mathrm{i}\phi}$, we obtain the needed relation:

$$\mathrm{e}^{\pm\mathrm{i}\phi}\left(\frac{\partial}{\partial\theta}\pm\mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\right) = r\cos\theta\Big(\frac{\partial}{\partial x}\pm\mathrm{i}\frac{\partial}{\partial y}\Big) - r\sin\theta\,\mathrm{e}^{\pm\mathrm{i}\phi}\frac{\partial}{\partial z}. \tag{7.50}$$

With this expression in hand we set to work on L_+. In the position representation it is

$$\begin{aligned}L_+ &= -\mathrm{i}\left(y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}\right)+\left(z\frac{\partial}{\partial x}-x\frac{\partial}{\partial z}\right)\\ &= z\Big(\frac{\partial}{\partial x}+\mathrm{i}\frac{\partial}{\partial y}\Big)-(x+\mathrm{i}y)\frac{\partial}{\partial z}\\ &= r\cos\theta\Big(\frac{\partial}{\partial x}+\mathrm{i}\frac{\partial}{\partial y}\Big)-r\sin\theta\,\mathrm{e}^{\mathrm{i}\phi}\frac{\partial}{\partial z},\end{aligned} \tag{7.51}$$

so with equation (7.50) we can write

$$L_+ = \mathrm{e}^{\mathrm{i}\phi}\Big(\frac{\partial}{\partial\theta}+\mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\Big). \tag{7.52a}$$

Similarly

$$\begin{aligned}L_- &= -\mathrm{i}\left(y\frac{\partial}{\partial z}-z\frac{\partial}{\partial y}\right)-\left(z\frac{\partial}{\partial x}-x\frac{\partial}{\partial z}\right)\\ &= -z\Big(\frac{\partial}{\partial x}-\mathrm{i}\frac{\partial}{\partial y}\Big)+(x-\mathrm{i}y)\frac{\partial}{\partial z}\\ &= -\left\{r\cos\theta\Big(\frac{\partial}{\partial x}-\mathrm{i}\frac{\partial}{\partial y}\Big)-r\sin\theta\,\mathrm{e}^{-\mathrm{i}\phi}\frac{\partial}{\partial z}\right\}\\ &= -\mathrm{e}^{-\mathrm{i}\phi}\Big(\frac{\partial}{\partial\theta}-\mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\Big).\end{aligned} \tag{7.52b}$$

The state $|l,l\rangle$ with the largest permissible value of m for given l must satisfy the equation $L_+|l,l\rangle = 0$. Using equations (7.45) and (7.52a), in the position representation this reads

$$\frac{\partial K_{ll}}{\partial\theta} - l\cot\theta K_{ll} = 0. \tag{7.53}$$

Table 7.1 The first six spherical harmonics

	m 0	 ± 1	 ± 2
Y_0^m	$\sqrt{\frac{1}{4\pi}}$		
Y_1^m	$\sqrt{\frac{6}{8\pi}}\cos\theta$	$\mp\sqrt{\frac{3}{8\pi}}\sin\theta\mathrm{e}^{\pm\mathrm{i}\phi}$	
Y_2^m	$\sqrt{\frac{10}{32\pi}}(3\cos^2\theta-1)$	$\mp\sqrt{\frac{15}{32\pi}}\sin 2\theta\mathrm{e}^{\pm\mathrm{i}\phi}$	$\sqrt{\frac{15}{32\pi}}\sin^2\theta\mathrm{e}^{\pm 2\mathrm{i}\phi}$

This is a first-order linear differential equation. Its integrating factor is $\exp\bigl(-l\int\mathrm{d}\theta\,\cot\theta\bigr)=\sin^{-l}\theta$, so its solution is $K_{ll}=R(r)\sin^l\theta$, where R is an arbitrary function. Substituting this form of K_{ll} into equation (7.45), we conclude that

$$\psi_{ll}(r,\theta,\phi)=R(r)\sin^l\theta\,\mathrm{e}^{\mathrm{i}l\phi}. \tag{7.54}$$

From equation (7.54) we can obtain the wavefunctions ψ_{lm} of states with smaller values of m simply by applying the differential operator L_-. For example

$$\begin{aligned}\psi_{l(l-1)}(r,\theta,\phi)&=\text{constant}\times R(r)\mathrm{e}^{-\mathrm{i}\phi}\Bigl(-\frac{\partial}{\partial\theta}+\mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\Bigr)\sin^l\theta\,\mathrm{e}^{\mathrm{i}l\phi}\\&=\text{constant}\times R(r)\sin^{l-1}\theta\cos\theta\mathrm{e}^{\mathrm{i}(l-1)\phi}.\end{aligned} \tag{7.55}$$

Hence, the eigenfunctions of L^2 and L_z for given l all have the same radial dependence, $R(r)$. The function of θ,ϕ that multiplies R in ψ_{lm} is conventionally denoted Y_l^m and called a **spherical harmonic**. The normalisation of Y_l^m is chosen such that

$$\int\mathrm{d}^2\Omega\,|\mathrm{Y}_l^m|^2=1\quad\text{with}\quad\mathrm{d}^2\Omega\equiv\sin\theta\,\mathrm{d}\theta\mathrm{d}\phi \tag{7.56}$$

the element of solid angle. We have shown that

$$\mathrm{Y}_l^l\propto\sin^l\theta\mathrm{e}^{\mathrm{i}l\phi}\quad\text{and}\quad\mathrm{Y}_l^{l-1}\propto\sin^{l-1}\theta\cos\theta\mathrm{e}^{\mathrm{i}(l-1)\phi}. \tag{7.57}$$

The normalising constants can be determined by first evaluating the integral

$$\int\mathrm{d}^2\Omega\,\sin^{2l}\theta=4\pi\,2^{2l}\frac{(l!)^2}{(2l+1)!} \tag{7.58}$$

involved in the normalisation of Y_l^l, and then dividing by the factor α_- of equation (7.47b) each time L_- is applied.

The spherical harmonics Y_l^m for $l\le 2$ are listed in Table 7.1. Figures 7.4 and 7.5 show contour plots of several spherical harmonics. Since spherical harmonics are functions on the unit sphere, the figures show a series of balls with contours drawn on them. Since spherical harmonics are complex functions we had to decide whether to show the real part, the imaginary part, the modulus or the phase of the function. We decided it was most instructive to plot contours on which the real part is constant; when the real part is positive, the contour is full, and when it is negative, the contour is dotted.

For large l, Y_l^l is significantly non-zero only where $\sin\theta\simeq 1$, i.e., around the equator, $\theta=\pi/2$ – the leftmost panel of Figure 7.4 illustrates this case. The first l applications of L_- each introduce a term that contains one less power of $\sin\theta$ and an extra power of $\cos\theta$. Consequently, as m diminishes from l to zero, the region of the sphere in which Y_l^m is significantly non-zero gradually spreads from the equator toward the poles – compare the leftmost and rightmost panels of Figure 7.4. These facts make good sense physically:

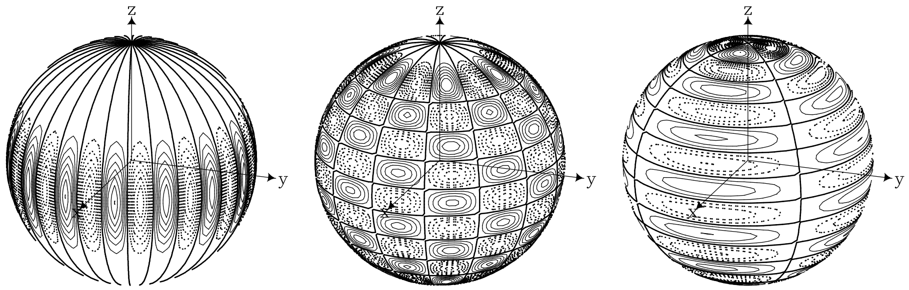


Figure 7.4 Contours of $\Re(\mathrm{Y}_{15}^m)$ on the unit sphere for $m = 15$ (left), $m = 7$ (centre) and $m = 2$ (right). The contours on which $\Re(\mathrm{Y}_{15}^m) = 0$ are the heavy curves, while contours on which $\Re(\mathrm{Y}_{15}^m) < 0$ are dotted. Contours of the imaginary part of Y_l^m would look the same except shifted in azimuth by half the distance between the heavy curves of constant azimuth.

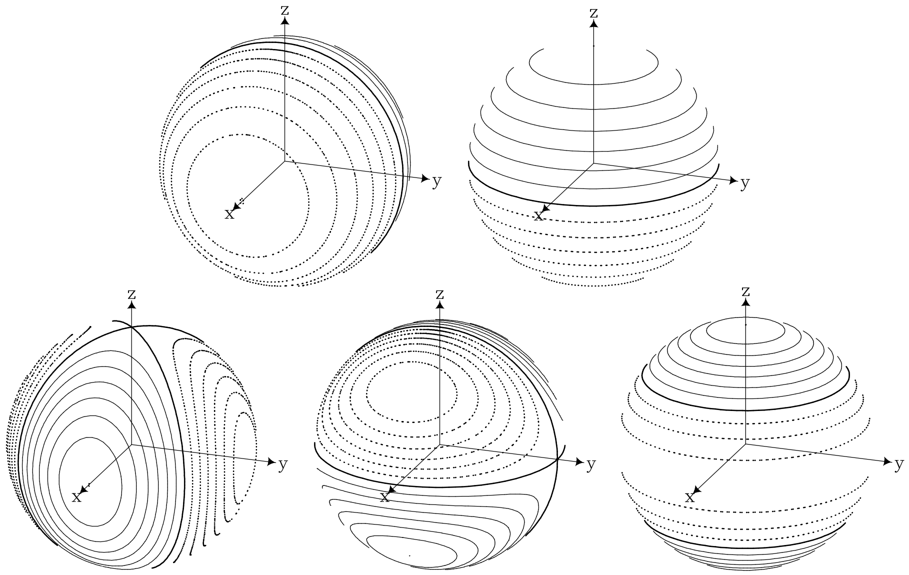


Figure 7.5 Top row: contours of $\Re(\mathrm{Y}_1^m)$ for $m = 1$ (left) and 0 (right) with line styles having the same meaning as in Figure 7.4. Contours of the imaginary part of Y_l^1 would look the same as the left panel but with the circles centred on the y axis. Bottom row: contours of $\Re(\mathrm{Y}_2^m)$ for $m = 2$ (left), $m = 1$ (centre) and $m = 0$ (right).

Y_l^l is the wavefunction of a particle that has essentially all its orbital angular momentum parallel to the z axis, so the particle should not stray far from the xy plane. Hence Y_l^l, the amplitude to find the particle at θ, should be small for θ significantly different from $\pi/2$. As m diminishes the orbital plane is becoming more inclined to the xy plane, so we are likely to find the particle further and further from the plane. This is why Y_l^m increases away from the equator as m decreases.

For large l the phase of Y_l^l changes rapidly with ϕ (leftmost panel of Figure 7.4). This is to be expected, because the particle's large orbital angular momentum, $l\hbar$, implies that the particle has a substantial tangential motion within the xy plane. From classical physics we estimate its tangential momentum at $p = l\hbar/r$, and from quantum mechanics we know that this implies that the wavefunction must change its phase at a rate $p/\hbar = l/r$ radians per unit distance. This estimate agrees precisely with the rate of change of phase with distance around the equator arising from the factor $\mathrm{e}^{\mathrm{i}l\phi}$ in Y_l^l. When m is significantly smaller than l (rightmost panel of Figure 7.4), the rate of change of the wavefunction's phase with increasing ϕ is smaller because the particle's tangential momentum is not all in the direction of increasing ϕ. Hence $\mathrm{Y}_l^m \propto \mathrm{e}^{\mathrm{i}m\phi}$.

For any value of m, L_x and L_y both have zero expectation values, as follows immediately from the relation $L_x = \frac{1}{2}(L_+ + L_-)$. So the orientation of the component of the angular momentum vector that lies in the xy plane is completely uncertain. Because of this uncertainty, the modulus of Y_l^m is independent of ϕ, so there is no trace of an inclined orbital plane when $m < l$. An orbital plane becomes defined if there is some uncertainty in L_z, with the result that there are non-zero amplitudes $\psi_m = \langle l, m|\psi\rangle$ for several values of m. In this case quantum interference between states of well-defined L_z can generate a peak in $|\langle\mathbf{x}|\psi\rangle|^2$ along a great circle that is inclined to the equator.

7.2.4 Orbital angular momentum and parity

In §4.1.4 we defined the parity operator P, which turns a state with wavefunction $\psi(\mathbf{x})$ into the state that has wavefunction $\psi'(\mathbf{x}) \equiv \psi(-\mathbf{x})$. We now show that wavefunctions that are proportional to a spherical harmonic Y_l^m are eigenfunctions of P with eigenvalue $(-1)^l$.

In polar coordinates the transformation $\mathbf{x} \to -\mathbf{x}$ is effected by $\theta \to \pi - \theta$, $\phi \to \phi + \pi$. Under this mapping, $\sin\theta = \sin(\pi - \theta)$ is unchanged, while $\mathrm{e}^{\mathrm{i}l\phi} \to \mathrm{e}^{\mathrm{i}l\pi}\mathrm{e}^{\mathrm{i}l\phi} = (-1)^l\mathrm{e}^{\mathrm{i}l\phi}$. By equation (7.57), $\mathrm{Y}_l^l \propto \sin^l\theta\mathrm{e}^{\mathrm{i}l\phi}$, so $\mathrm{Y}_l^l \to (-1)^l\mathrm{Y}_l^l$. That is, Y_l^l has even parity if l is an even number and odd parity otherwise.

In §4.1.4 we saw that $\mathbf{x}$ and $\mathbf{p}$ are odd-parity operators: $P\mathbf{x} = -\mathbf{x}P$. From this and the fact that the orbital angular momentum operators L_i are sums of products of a component of $\mathbf{x}$ and a component of $\mathbf{p}$, it follows that both the L_i and the ladder operators $L_\pm = L_x \pm \mathrm{i}L_y$ are even-parity operators. Now $\mathrm{Y}_l^{m-1} = L_-\mathrm{Y}_l^m/\alpha_-$, where α_- is a constant, so applying the parity operator

$$P\mathrm{Y}_l^{l-1} = \frac{1}{\alpha_-}L_-P\mathrm{Y}_l^l = (-1)^l\frac{1}{\alpha_-}L_-\mathrm{Y}_l^l = (-1)^l\mathrm{Y}_l^{l-1}. \tag{7.59}$$

That is, Y_l^{l-1} has the same parity as Y_l^l. Since all the Y_l^m for a given l can be obtained by repeated application of L_- to Y_l^l, it follows that they all have the same parity, $(-1)^l$.

7.2.5 Orbital angular momentum and kinetic energy

We now derive a very useful decomposition of the kinetic energy operator $H_\mathrm{K} \equiv p^2/2m$ into a sum of operators for the radial and tangential kinetic energies. First we show that L^2 is intimately related to the Laplacian operator ∇^2. From the definition (7.46) of the ladder operators for orbital angular momentum, we have

$$\begin{aligned} L_+L_- &= (L_x + \mathrm{i}L_y)(L_x - \mathrm{i}L_y) = L_x^2 + L_y^2 + \mathrm{i}[L_y, L_x] \\ &= L_x^2 + L_y^2 + L_z. \end{aligned} \tag{7.60}$$

Hence with equations (7.52) we may write

$$\begin{aligned} L^2 &= L_+L_- - L_z + L_z^2 \\ &= \mathrm{e}^{\mathrm{i}\phi}\Big(\frac{\partial}{\partial\theta} + \mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\Big)\Big\{\mathrm{e}^{-\mathrm{i}\phi}\Big(-\frac{\partial}{\partial\theta} + \mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\Big)\Big\} + \mathrm{i}\frac{\partial}{\partial\phi} - \frac{\partial^2}{\partial\phi^2}. \end{aligned}$$

Differentiating out the right side

$$L^2 = -\frac{\partial^2}{\partial\theta^2} - \cot^2\theta\frac{\partial^2}{\partial\phi^2} + \cot\theta\Big(-\frac{\partial}{\partial\theta} + \mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\Big) - \mathrm{i}\csc^2\theta\frac{\partial}{\partial\phi} + \mathrm{i}\frac{\partial}{\partial\phi} - \frac{\partial^2}{\partial\phi^2}.$$

The first-order terms in $\partial/\partial\phi$ cancel because $\cot^2\theta - \csc^2\theta = -1$. This identity also enables us to combine the double derivatives in ϕ. Finally the single and double derivatives in θ can be combined so that the equation becomes

$$L^2 = -\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right\}, \tag{7.61}$$

which we recognise as $-r^2$ times the angular part of the Laplacian operator ∇^2.

Now we ask "what is the operator associated with radial momentum?". The obvious candidate is $\hat{\mathbf{r}}\cdot\mathbf{p}$, where $\hat{\mathbf{r}}$ is the unit vector in the radial direction. Unfortunately this operator is not Hermitian:

$$(\hat{\mathbf{r}}\cdot\mathbf{p})^\dagger = \mathbf{p}\cdot\hat{\mathbf{r}} \neq \hat{\mathbf{r}}\cdot\mathbf{p}, \tag{7.62}$$

so it is not an observable. This is a particular case of a general phenomenon: the product AB of two non-commuting observables A and B is never Hermitian. But it is easy to see that $\frac{1}{2}(AB+BA)$ *is* Hermitian. So we define

$$p_r \equiv \tfrac{1}{2}(\hat{\mathbf{r}}\cdot\mathbf{p} + \mathbf{p}\cdot\hat{\mathbf{r}}), \tag{7.63}$$

which is manifestly Hermitian. We will need an expression for p_r in the position representation. Replacing $\mathbf{p}$ by $-\mathrm{i}\hbar\boldsymbol{\nabla}$ we have

$$p_r = -\frac{\mathrm{i}\hbar}{2}\left(\frac{1}{r}\mathbf{r}\cdot\boldsymbol{\nabla} + \boldsymbol{\nabla}\cdot(\mathbf{r}/r)\right). \tag{7.64}$$

From the chain rule it is straightforward to show that

$$r\frac{\partial}{\partial r} = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z} = \mathbf{r}\cdot\boldsymbol{\nabla}\ . \tag{7.65}$$

Moreover, $\boldsymbol{\nabla}\cdot\mathbf{r} = 3$, so equation (7.64) can be rewritten

$$\begin{aligned} p_r &= -\frac{\mathrm{i}\hbar}{2}\left(\frac{\partial}{\partial r} + \frac{3}{r} - \frac{r}{r^2} + \frac{\partial}{\partial r}\right) \\ &= -\mathrm{i}\hbar\left(\frac{\partial}{\partial r} + \frac{1}{r}\right). \end{aligned} \tag{7.66}$$

This expression enables us to find the commutator

$$[r, p_r] = -\mathrm{i}\hbar\left[r, \frac{\partial}{\partial r}\right] = \mathrm{i}\hbar. \tag{7.67}$$

Squaring both sides of equation (7.66) yields

$$\begin{aligned} p_r^2 &= -\hbar^2\left(\frac{\partial}{\partial r} + \frac{1}{r}\right)\left(\frac{\partial}{\partial r} + \frac{1}{r}\right) = -\hbar^2\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{1}{r^2} + \frac{1}{r^2}\right) \\ &= -\frac{\hbar^2}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right). \end{aligned} \tag{7.68}$$

We recognise this operator as $-\hbar^2$ times the radial part of the Laplacian operator ∇^2. Since we have shown that L^2 is $-r^2$ times the angular part of the Laplacian (eq. 7.61), it follows that $\nabla^2 = -(p_r^2/\hbar^2 + L^2/r^2)$. Consequently, the kinetic-energy operator $H_\mathrm{K} = p^2/2m = -(\hbar^2/2m)\nabla^2$ can be written

$$H_\mathrm{K} = \frac{1}{2m}\left(p_r^2 + \frac{\hbar^2 L^2}{r^2}\right). \tag{7.69}$$

The physical interpretation of this equation is clear: classically, the orbital angular momentum $\hbar\mathbf{L}$ is $m\mathbf{r}\times\mathbf{v} = mrv_t$, where v_t is the tangential speed, so the term $\hbar^2L^2/2mr^2 = \frac{1}{2}mv_t^2$ is the kinetic energy associated with the tangential motion. On the other hand $p_r^2/2m = \frac{1}{2}mv_r^2$, so this term represents the kinetic energy due to radial motion, as we would expect. For future reference we note that the kinetic-energy operator can be also written

$$H_\mathrm{K} = -\frac{\hbar^2}{2m}\left\{\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) - \frac{L^2}{r^2}\right\}. \tag{7.70}$$

Table 7.2 The first five Legendre polynomials

	l				
	0	1	2	3	4
$P_l(\mu)$	1	μ	$\frac{1}{2}(3\mu^2-1)$	$\frac{1}{2}(5\mu^3-3\mu)$	$\frac{1}{8}(35\mu^4-30\mu^2+3)$

7.2.6 Legendre polynomials

The spherical harmonic Y_l^0 is special in that it is a function of θ only. We now show that it is, in fact, a polynomial in $\cos\theta$. In the interval $0 \le \theta \le \pi$ of interest, θ is a monotone function of $\mu \equiv \cos\theta$, so without any loss of generality we may take Y_l^0 to be proportional to a function $P_l(\mu)$. On this understanding, P_l is an eigenfunction of L^2 with eigenvalue $l(l+1)$. Transforming the independent variable from θ to μ in our expression (7.61) for L^2, we find that P_l must satisfy **Legendre's equation**:

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left((1-\mu^2)\frac{\mathrm{d}P_l}{\mathrm{d}\mu}\right)+l(l+1)P_l=0. \tag{7.71}$$

We look for polynomial solutions of this equation. Putting in the trial solution $P_l=\sum_n b_n\mu^n$, we find

$$\sum_n b_n\left\{n(n-1)\mu^{n-2}-n(n+1)\mu^n+l(l+1)\mu^n\right\}=0. \tag{7.72}$$

This equation must be valid for any value of μ in the interval $(-1,1)$, which will be possible only if the coefficient of each and every power of μ individually vanishes. The coefficient of μ^k is

$$0=b_{k+2}(k+2)(k+1)-b_k\left\{k(k+1)-l(l+1)\right\}. \tag{7.73}$$

For $k=0$ the expression connects b_2 to b_0, while for $k=2$ it relates b_4 to b_2, and so on. Thus from this equation we can express b_n as a multiple of b_0 for even n, and as a multiple of b_1 for odd n. Moreover, if l is an even number, we know from our discussion of parity that P_l must be an even function of μ, so in this case b_n must vanish for n odd. Finally, b_n will vanish for n even and greater than l on account of the vanishing of the curly bracket in equation (7.73) when $k=l$. This completes the proof that for even l, $P_l(\mu)$ is a polynomial or order l. An extremely similar argument shows that $P_l(\mu)$ is also a polynomial of order l when l is odd. The first five **Legendre polynomials** are listed in Table 7.2.

The conventional normalisation of the Legendre polynomial P_l is the requirement that $P_l(1)=1$. With this property, the P_l are not orthonormal. In fact

$$\int_{-1}^{1}\mathrm{d}\mu\,P_l(\mu)P_{l'}(\mu)=\frac{2}{2l+1}\delta_{ll'}. \tag{7.74}$$

From this result it easily follows that the proportionality constant between $P_l(\cos\theta)$ and the orthonormal functions $\mathrm{Y}_l^0(\theta)$ is such that

$$\mathrm{Y}_l^0(\theta)=\sqrt{\frac{2l+1}{4\pi}}\,P_l(\cos\theta). \tag{7.75}$$

7.3 Three-dimensional harmonic oscillator

In this section we discuss the dynamics of a particle that moves in three dimensions subject to a central force that is proportional to the particle's distance from the origin. So the Hamiltonian is

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 r^2. \tag{7.76}$$

If we use Cartesian coordinates, this Hamiltonian becomes the sum of three copies of the Hamiltonian of the one-dimensional harmonic oscillator that was the subject of §3.1:

$$H = H_x + H_y + H_z, \tag{7.77}$$

where, for example, $H_x = (p_x^2/2m) + \frac{1}{2}m\omega^2 x^2$. These one-dimensional Hamiltonians commute with one another. So there is a complete set of mutual eigenkets. Let $|n_x, n_y, n_z\rangle$ be the state that is an eigenket of H_x with eigenvalue $(n_x + \frac{1}{2})\hbar\omega$ eq. 3.11, etc. Then $|n_x, n_y, n_z\rangle$ will be an eigenket of the three-dimensional Hamiltonian (7.76) with eigenvalue

$$E = (n_x + n_y + n_z + \tfrac{3}{2})\hbar\omega. \tag{7.78}$$

Moreover, in the position representation the wavefunction of this state is just a product of three of the wavefunctions we derived for stationary states of a one-dimensional oscillator

$$\psi(\mathbf{x}) = u_{n_x}(x)u_{n_y}(y)u_{n_z}(z). \tag{7.79}$$

In view of these considerations it might be thought that there is nothing we do not know about the Hamiltonian (7.76). However, it is instructive to reanalyse the system from a more physical point of view, that recognises that the system is spherically symmetric. We have seen that $[L_i, p^2] = 0$, and $[L_i, r^2] = 0$, so $[L_i, H] = 0$ and $[L^2, H] = 0$. From this result it follows that there is a complete set of mutual eigenstates of H, L^2 and L_z. Very few of the eigenstates obtained from the one-dimensional Hamiltonians are eigenstates of either L^2 or L_z. We now show how the eigenvalue problem associated with (7.76) can be solved in a way that yields mutual eigenkets of H, L^2 and L_z. This exercise is instructive in itself, and some technology that we will develop along the way will prove extremely useful when we analyse the hydrogen atom in Chapter 8.

We use equation (7.69) to eliminate p^2 from equation (7.76)

$$H = \frac{p_r^2}{2m} + \frac{\hbar^2 L^2}{2mr^2} + \tfrac{1}{2}m\omega^2 r^2. \tag{7.80}$$

We can assume that our energy eigenstates are also eigenstates of L^2, so in this Hamiltonian we can replace L^2 by an eigenvalue $l(l+1)$. Hence we wish to find the eigenvalues of the **radial Hamiltonian**

$$H_l = \frac{p_r^2}{2m} + \frac{l(l+1)\hbar^2}{2mr^2} + \tfrac{1}{2}m\omega^2 r^2. \tag{7.81}$$

Our determination of the allowed energies of a one-dimensional harmonic oscillator exploited the dimensionless operators A and $A^\dagger$, which rather nearly factorise $H/\hbar\omega$. So here we define the operator

$$A_l \equiv \frac{1}{\sqrt{2m\hbar\omega}}\left(\mathrm{i}p_r - \frac{(l+1)\hbar}{r} + m\omega r\right). \tag{7.82}$$

The product of A and its Hermitian adjoint $A_l^\dagger$ is

$$\begin{aligned} A_l^\dagger A_l &= \frac{1}{2m\hbar\omega}\left(-\mathrm{i}p_r - \frac{(l+1)\hbar}{r} + m\omega r\right)\left(\mathrm{i}p_r - \frac{(l+1)\hbar}{r} + m\omega r\right) \\ &= \frac{1}{2m\hbar\omega}\left\{p_r^2 + \left(-\frac{(l+1)\hbar}{r} + m\omega r\right)^2 + \mathrm{i}\left[-\frac{(l+1)\hbar}{r} + m\omega r, p_r\right]\right\} \\ &= \frac{1}{2m\hbar\omega}\left\{p_r^2 + \frac{l(l+1)\hbar^2}{r^2} + m^2\omega^2 r^2 - (2l+3)\hbar m\omega\right\}. \end{aligned} \tag{7.83}$$

Comparing the right side with equation (7.81), we see that

$$H_l = \hbar\omega\left\{A_l^\dagger A_l + (l + \tfrac{3}{2})\right\}, \tag{7.84}$$

which bears a strong similarity to equation (3.3) for the one-dimensional harmonic oscillator.

The commutator of A_l and $A_l^\dagger$ is

$$\begin{aligned} [A_l, A_l^\dagger] &= \frac{1}{2m\hbar\omega}\left[\left(\mathrm{i}p_r - \frac{(l+1)\hbar}{r} + m\omega r\right), \left(-\mathrm{i}p_r - \frac{(l+1)\hbar}{r} + m\omega r\right)\right] \\ &= \frac{\mathrm{i}}{m\hbar\omega}\left[p_r, \left(-\frac{(l+1)\hbar}{r} + m\omega r\right)\right] \\ &= \frac{(l+1)\hbar}{m\omega r^2} + 1, \end{aligned} \tag{7.85}$$

where we have used equation (7.67) to reach the last line. This result can be written more usefully in the form

$$[A_l, A_l^\dagger] = \frac{H_{l+1} - H_l}{\hbar\omega} + 1. \tag{7.86}$$

From this expression and equation (7.84) we can easily calculate the commutator of H_l with A_l:

$$[A_l, H_l] = \hbar\omega[A_l, A_l^\dagger A_l] = \hbar\omega[A_l, A_l^\dagger]A_l = (H_{l+1} - H_l + \hbar\omega)A_l. \tag{7.87}$$

Now let $|E, l\rangle$ be an eigenket of H_l with eigenvalue E:

$$H_l|E, l\rangle = E|E, l\rangle. \tag{7.88}$$

We multiply both sides of the equation by A_l and use equation (7.87) to reverse the order of A_l and H_l:

$$\begin{aligned} EA_l|E, l\rangle = A_l H_l|E, l\rangle &= (H_l A_l + [A_l, H_l])|E, l\rangle \\ &= (H_{l+1} + \hbar\omega)A_l|E, l\rangle. \end{aligned} \tag{7.89}$$

On rearrangement this yields

$$H_{l+1}(A_l|E, l\rangle) = (E - \hbar\omega)(A_l|E, l\rangle), \tag{7.90}$$

which says that $A_l|E, l\rangle$ is an eigenket of H_{l+1} for the eigenvalue $E - \hbar\omega$,

$$A_l|E, l\rangle = \alpha_-|E - \hbar\omega, l + 1\rangle, \tag{7.91}$$

Where α_- is a normalising constant.

A_l creates the radial wavefunction for a state that has *more* orbital angular momentum and *less* energy than the state with which it started. That is, A_l diminishes the radial kinetic energy by some amount and adds a

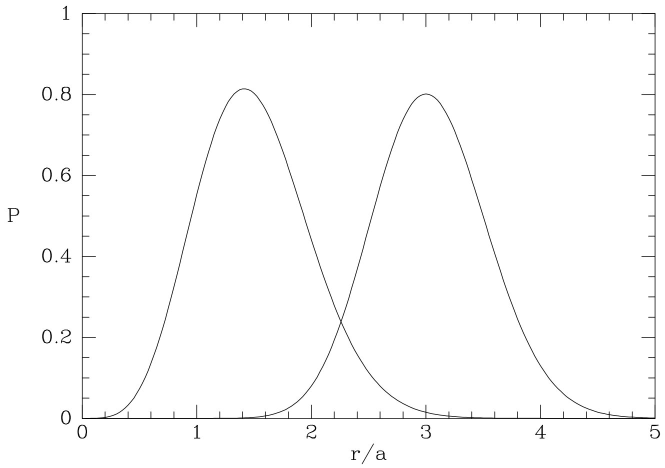


Figure 7.6 Radial probability distributions of circular orbits in the three-dimensional harmonic oscillator potential for $l = 1$ and $l = 8$. The scale radius a is defined by equation (3.14).

smaller quantity of energy to the tangential motion. If we repeat this process a sufficient number of times, by following A_l with A_{l+1} and A_{l+1} with A_{l+2}, and so on, there will come a point at which no radial kinetic energy remains – we will have reached the quantum equivalent of a circular orbit. The next application of A_l must annihilate the wavefunction. Hence $A_l|E,\mathcal{L}\rangle = 0$, where $\mathcal{L}(E)$ is the largest allowed value of l for energy E. If we operate on $|E,\mathcal{L}\rangle$ with H_l, we find with equation (7.84) that

$$E|E,\mathcal{L}\rangle = H_l|E,\mathcal{L}\rangle = \hbar\omega(\mathcal{L}+\tfrac{3}{2})|E,\mathcal{L}\rangle, \tag{7.92}$$

so

$$E = (\mathcal{L}+\tfrac{3}{2})\hbar\omega \quad \text{and} \quad \mathcal{L}(E) = \frac{E}{\hbar\omega} - \tfrac{3}{2}.$$

Since $\mathcal{L}$ is a non-negative integer, it follows that the ground-state energy is $\frac{3}{2}\hbar\omega$ and that the ground state has no angular momentum. In general $E/\hbar\omega$ is any integer plus $\frac{3}{2}$. These values of the allowed energies agree perfectly with what we could have deduced by treating H as a sum of three one-dimensional harmonic-oscillator Hamiltonians.

We shall define a **circular orbit** to be one that has the maximum angular momentum possible at its energy. We obtain the radial wavefunctions of these by writing the equation $A_l|E,\mathcal{L}\rangle = 0$ in the position representation. With equations (7.82) and (7.66) this equation becomes

$$\left(\frac{\partial}{\partial r} + \frac{1}{r} - \frac{l+1}{r} + \frac{m\omega}{\hbar} r\right) u_{\mathcal{L}}(r) = 0. \tag{7.93}$$

This is a first-order linear differential equation. Its integrating factor is

$$\exp\left\{\int \mathrm{d}r \left(-\frac{l}{r} + \frac{m\omega}{\hbar} r\right)\right\} = r^{-l}\exp\left(\frac{m\omega}{2\hbar} r^2\right), \tag{7.94}$$

so the solution of equation (7.93) is

$$u_{\mathcal{L}}(r) = \text{constant} \times r^l \exp\left(-\frac{m\omega}{2\hbar} r^2\right). \tag{7.95}$$

Notice that the exponential factor is simply the product of three exponential factors from equation (3.14), one in x, one in y and one in z. The wavefunction varies with r, so a circular orbit does have some radial kinetic energy. In the limit of large l in which classical mechanics applies, the radial kinetic energy is negligible compared to the tangential kinetic energy, and we neglect it. But it never really vanishes.

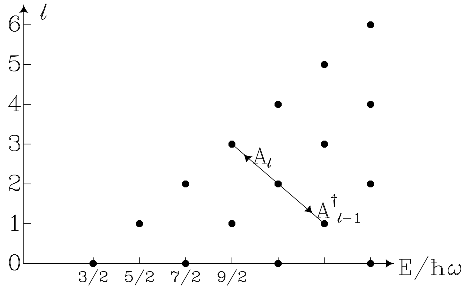


Figure 7.7 The (E, l) plane and the action of A_l and $A^\dagger_{l-1}$.

Equation (7.95) gives the radial wavefunction for a circular orbit. The complete wavefunction is $\psi(\mathbf{x}) = u_{\mathcal{L}}(r)\mathrm{Y}_l^m(\theta,\phi)$, and since $\int \mathrm{d}^2\Omega\, \mathrm{Y}_l^m = 1$, the radial probability density is $P(r) = r^2 u_{\mathcal{L}}^2 \propto r^{2l+2}\mathrm{e}^{-r^2/a^2}$, where the factor r^2 arises from the expression for the volume element $\mathrm{d}^3\mathbf{x}$ in spherical polar coordinates, and $a = \sqrt{\hbar/m\omega}$. This density is plotted in Figure 7.6 for $l = 1$ and $l = 8$. For $r < \sqrt{l+1}\,a$, P rises as r^{2l+2}. At larger radii it falls rapidly as the Gaussian factor takes over. Hence the uncertainty in r is $\sim a$, which is a small fraction of r when l is not small.

We may obtain the radial wavefunctions of more eccentric orbits by showing that $A_l^\dagger$ is a raising operator. Equation (7.86) implies that

$$[A_{l-1}, A^\dagger_{l-1}] = \frac{H_l - H_{l-1}}{\hbar\omega} + 1. \tag{7.96}$$

Commuting both sides with $A^\dagger_{l-1}$ yields

$$[H_l, A^\dagger_{l-1}] = [H_{l-1}, A^\dagger_{l-1}] + \hbar\omega\big[[A_{l-1}, A^\dagger_{l-1}], A^\dagger_{l-1}\big]. \tag{7.97}$$

When we use equation (7.84) to eliminate H_{l-1} and expand the inner commutator on the right as a difference of products, we obtain an equation analogous to (7.87):

$$[H_l, A^\dagger_{l-1}] = \hbar\omega[A_{l-1}, A^\dagger_{l-1}]A^\dagger_{l-1} = (H_l - H_{l-1} + \hbar\omega)A^\dagger_{l-1}, \tag{7.98}$$

where the second equality exploits equation (7.86). We are now ready to multiply both sides of equation (7.88) by $A^\dagger_{l-1}$. We find

$$EA^\dagger_{l-1}|E,l\rangle = (H_l A^\dagger_{l-1} + [A^\dagger_{l-1}, H_l])|E,l\rangle = (H_{l-1} - \hbar\omega)A^\dagger_{l-1}|E,l\rangle, \tag{7.99}$$

so

$$H_{l-1}(A^\dagger_{l-1}|E,l\rangle) = (E + \hbar\omega)(A^\dagger_{l-1}|E,l\rangle). \tag{7.100}$$

Thus, we have shown that

$$A^\dagger_{l-1}|E,l\rangle = \alpha_+|E + \hbar\omega, l-1\rangle, \tag{7.101}$$

where α_+ is a normalising constant. By writing $A^\dagger_{l-1}$ in the position representation, we can generate the wavefunctions of all non-circular orbits by repeatedly applying $A^\dagger_{l-1}$ to the wavefunction of an appropriate circular orbit. We start with the product of r^l and a Gaussian factor [equation (7.95)]. From this the first application of $A^\dagger_{l-1}$ generates terms proportional to r^{l+1} and r^{l-1} times the Gaussian (Problem 7.21). The next application generates three terms, r^{l+2}, r^l and r^{l-2} times the Gaussian, and so on. Consequently the number of radial **nodes** – radii at which the wavefunction vanishes – increases by one with each application of $A^\dagger_{l-1}$, and the wavefunction oscillates more and more rapidly in radius as $A^\dagger_{l-1}$ invests a larger and larger fraction of the particle's kinetic energy in radial motion.

Figure 7.7 helps to organise the generation of radial wavefunctions. Each dot represents a radial wavefunction. From the dot at (E, l), operating with A_l carries one to the next dot up and to the left, while operating with $A^\dagger_{l-1}$ carries one to the next dot down and to the right. At half the energies only even values of l occur, and only odd values of l occur at the other half of the energies. In Problem 7.18 you can show that, when one bears in mind that each dot gives rise to $2l+1$ complete wavefunctions, the number of wavefunction with energy E that we obtain in this way agrees with the number that we would obtain using wavefunctions of the one-dimensional harmonic oscillator via equation (7.79).

7.4 Spin angular momentum

In §7.2.1 we saw that the difference between **J** and **L** is that **J** is the generator for complete rotations of the system, while **L** is the generator for displacements of the system around circles, while leaving its orientation fixed (Figure 7.3). Consequently the difference

$$\mathbf{S} \equiv \mathbf{J} - \mathbf{L} \tag{7.102}$$

is the generator for changes of orientation that are not accompanied by any motion of the system as a whole. Since **J** and **L** are vector operators, **S** is also a vector operator. Its components are called the **spin operators**.

We saw in §7.2 that **L** has exactly the same commutation relations as **J** with any function of the position and momentum operators only. From this fact and the definition (7.102), it follows that **S** commutes with all such functions. In particular $[\mathbf{S}, \mathbf{x}] = [\mathbf{S}, \mathbf{p}] = [\mathbf{S}, \mathbf{L}] = 0$. This essentially tells us that **S** has nothing to do with a system's location, nor the way in which it may or may not be moving. **S** is associated with *intrinsic* properties of the system.

The components S_i of the spin operator inherit the usual angular momentum commutation rules from J_i and L_i:

$$\begin{aligned}[S_i, S_j] &= [J_i - L_i, J_j - L_j] \\ &= [J_i, J_j] - [L_i, J_j] - [J_i, L_j] + [L_i, L_j] \\ &= \mathrm{i}\sum_k \epsilon_{ijk}(J_k - L_k - L_k + L_k) \\ &= \mathrm{i}\sum_k \epsilon_{ijk} S_k.\end{aligned} \tag{7.103}$$

We define $S^2 \equiv \mathbf{S}\cdot\mathbf{S}$ and then equation (7.103) ensures that

$$[\mathbf{S}, S^2] = 0. \tag{7.104}$$

Because the S_i have exactly the same form of commutation relations as the J_i, we know that the possible eigenvalues of S^2 are the numbers $0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$ and that for given s the eigenvalues m of the S_i move in integer steps from $-s$ to s. Can s take half-integer values? This question is answered affirmatively by equation (7.102); since $[J_z, L_z] = 0$ we can find a complete set of states that simultaneously have well-defined values of both J_z and L_z. In general, the J_z eigenvalue could be either an integer or half-integer, whereas the L_z eigenvalue must be an integer. The difference $S_z = J_z - L_z$ must then be either an integer or half-integer.

In the rest of this book we will make extensive use of commutation relations involving angular momentum operators. In Table 7.3 these have been gathered for later reference.

Table 7.3 Commutators involving angular momentum

$[J_i, v_j] = \mathrm{i}\sum_k \epsilon_{ijk} v_k$	where **v** is any vector or pseudovector
$[J_i, s] = 0$	where s is any scalar or pseudoscalar
$[L_i, w_j] = \mathrm{i}\sum_k \epsilon_{ijk} w_k$	where **w** is any vector or pseudovector function of only **x**, **p** and constant scalars and pseudoscalars
$[L_i, f] = 0$	where f is any scalar or pseudoscalar function of only **x**, **p** and constant scalars and pseudoscalars
$[S_i, w] = 0$	where w is any function of spatial operators

The following are important special cases of the above results

$[J_i, J_j] = \mathrm{i}\sum_k \epsilon_{ijk} J_k$	$[J_i, L_j] = \mathrm{i}\sum_k \epsilon_{ijk} L_k$	$[J_i, S_j] = \mathrm{i}\sum_k \epsilon_{ijk} S_k$
$[L_i, L_j] = \mathrm{i}\sum_k \epsilon_{ijk} L_k$	$[S_i, S_j] = \mathrm{i}\sum_k \epsilon_{ijk} S_k$	$[L_i, S_j] = 0$
$[J_i, J^2] = 0$	$[J_i, L^2] = 0$	$[J_i, S^2] = 0$
$[L_i, L^2] = 0$	$[L_i, S^2] = 0$	$[S_i, L^2] = 0$
$[S_i, S^2] = 0$	$[L^2, J^2] = 0$	$[S^2, J^2] = 0$

Since $\mathbf{J} = \mathbf{L} + \mathbf{S}$ and therefore $J^2 = L^2 + S^2 + 2\mathbf{L}\cdot\mathbf{S}$ we also have

$$[L_i, J^2] = 2\mathrm{i}\sum_{jk} \epsilon_{ijk} S_j L_k \qquad [S_i, J^2] = 2\mathrm{i}\sum_{jk} \epsilon_{ijk} L_j S_k$$

7.4.1 Spin and orientation

We have several times stated without proof that the orientation of the system is encoded in the amplitudes a_{jm} for the system to be found in states of well defined angular momentum, $|j, m\rangle$. We now begin to justify this claim. For simplicity we consider spin angular momentum because we want to focus on the orientation of our system without concerning ourselves with its location. However, what we refer to as 'spin' is the total intrinsic angular momentum of the system. If the latter is a hydrogen atom, for example, it may contain a contribution from the orbital angular momentum of the electron in addition to the contributions from the intrinsic spins of the electron and the proton.

Since the S_i are Hermitian operators, any state $|\psi\rangle$ may be expanded in terms of the complete set of eigenstates $|s, m\rangle$ of, say, S_z and S^2. We have seen that these states are labelled by an integer or half integer s, with $-s \le m \le s$, so the complete expansion is

$$|\psi\rangle = \sum_{s=0,\frac{1}{2},1,\dots} \sum_{m=-s}^{s} \langle s, m|\psi\rangle |s, m\rangle. \tag{7.105}$$

Fortunately, systems for which quantum mechanical effects are significant rarely have more than a handful of non-zero amplitudes $\langle s, m|\psi\rangle$ in the sum of equation (7.105). In the simplest case we have an object with $s = 0$, a **spin-zero** object, such a pion. The sum in equation (7.105) contains only one eigenstate with $s = 0$, the state $|0,0\rangle$ because an object with no spin cannot have any spin angular momentum around the z axis.

When we rotate an object about the direction $\boldsymbol{\alpha}$ without translating it, its state is updated by the operator $U(\boldsymbol{\alpha}) = \mathrm{e}^{-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{S}}$ (cf. equation 4.13). When we apply this operator to a state of a spin-zero object, the state emerges unchanged:

$$U(\boldsymbol{\alpha})|0,0\rangle = \exp\left(-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{S}\right)|0,0\rangle = |0,0\rangle. \tag{7.106}$$

Hence, a spin-zero object, like a perfect sphere, is completely unchanged by an arbitrary rotation. In view of their invariance under rotations, spin zero particles are sometimes known as **scalar particles**.[4]

Some very important systems require just the two terms in equation (7.105) that are associated with $s = \frac{1}{2}$. These systems are called **spin-half** objects. Electrons, quarks, protons and neutrons fall into this category. For example, by equation (7.105) the state of an electron can be written

$$|\mathrm{e}^-\rangle = \sum_{m=-\frac{1}{2}}^{\frac{1}{2}} \langle \tfrac{1}{2}, m|\mathrm{e}^-\rangle|\tfrac{1}{2}, m\rangle. \tag{7.107}$$

Because there are only two terms in this expansion, the quantum uncertainty in the orientation of a spin-half system is very great. We shall see that the most precise information we can have is that the end of the system's angular momentum vector lies on a small circle in a given hemisphere – for example, we could state that it lies within the northern rather than the southern hemisphere, or the western rather than the eastern hemisphere. Where it lies on the chosen small circle is shrouded in quantum uncertainty.

Another important class of systems contains those that have total spin quantum number $s = 1$. These systems are called **spin-one** objects. The W and Z bosons fall in this class. For a spin-one system, the expansion (7.105) reduces to just $(2s + 1) = 3$ terms. For example, the state of a Z boson can be written

$$|\mathrm{Z}\rangle = \sum_{m=-1}^{1} \langle 1, m|\mathrm{Z}\rangle|1, m\rangle. \tag{7.108}$$

We will see that we can constrain the end of the angular-momentum vector of a spin-one system to lie on a small circle within a chosen hemisphere, or on the equator that divides the two hemispheres.

The larger a system's spin s, the more precisely we can constrain the end of its angular momentum vector. It is rather as if systems were subject to random torques of a certain magnitude, and the faster it is spinning, the more stable its orientation can be in the face of the random torques. The same physical principle underlies the use of rifling in guns to stabilise the orientation of the projectile by imparting angular momentum to it as it flies down the barrel. A few concrete examples will clarify the physical interpretation of the quantum states $|s, m\rangle$.

7.4.2 Spin-half systems

As in equation (7.107), the state of any spin-half system may be expanded in terms of just two S_z eigenstates $|\frac{1}{2}, +\frac{1}{2}\rangle$ and $|\frac{1}{2}, -\frac{1}{2}\rangle$ which we will call

---

[4] Since both $\mathbf{J}$ and $\mathbf{S}$ commute with the parity operator P, behaviour under rotations does not tell us about behaviour under reflection, so spin zero particles could also be pseudoscalars. In fact, pions are pseudoscalar particles.

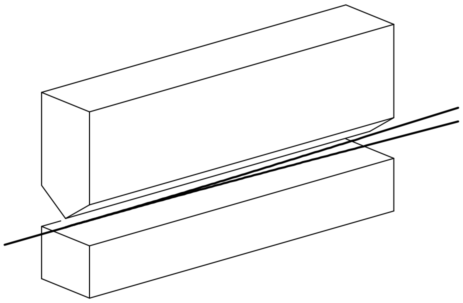


Figure 7.8 Schematic of a Stern–Gerlach filter. The atomic beam enters from the left. Between the pole pieces the magnetic field increases in intensity upwards, so atoms that have their spins aligned with **B** are deflected upwards and the other atoms are deflected downwards.

$|+\rangle$ and $|-\rangle$ respectively. Equation (7.107) then reads $|\psi\rangle = a|+\rangle + b|-\rangle$. In this basis we can write the operators as (cf. equation 2.16)[5]

$$S_x = \begin{pmatrix} \langle +|S_x|+\rangle & \langle +|S_x|-\rangle \\ \langle -|S_x|+\rangle & \langle -|S_x|-\rangle \end{pmatrix} \quad ; \quad S_y = \begin{pmatrix} \langle +|S_y|+\rangle & \langle +|S_y|-\rangle \\ \langle -|S_y|+\rangle & \langle -|S_y|-\rangle \end{pmatrix}$$

$$S_z = \begin{pmatrix} \langle +|S_z|+\rangle & \langle +|S_z|-\rangle \\ \langle -|S_z|+\rangle & \langle -|S_z|-\rangle \end{pmatrix}. \tag{7.109}$$

The elements of the matrix S_z are trivially evaluated because $|\pm\rangle$ are the eigenkets of S_z with eigenvalues $\pm\frac{1}{2}$. To evaluate the other two matrices we notice that $S_x = \frac{1}{2}(S_+ + S_-)$, and $S_y = \frac{1}{2\mathrm{i}}(S_+ - S_-)$, then use the relations $S_+|-\rangle = |+\rangle$ and $S_-|+\rangle = |-\rangle$ which follow from equations (7.5) and (7.7) for the spin operator. The result of these operations is

$$S_x = \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \; ; \; S_y = \tfrac{1}{2}\begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \; ; \; S_z = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{7.110}$$

The matrices appearing here are the **Pauli matrices**,

$$\sigma_x \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \; ; \; \sigma_y \equiv \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \; ; \; \sigma_z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \, , \tag{7.111}$$

so we can write $\mathbf{S} = \frac{1}{2}\boldsymbol{\sigma}$. It is straightforward to verify that the square of any Pauli matrix is the identity matrix:

$$\sigma_i^2 = I. \tag{7.112}$$

This result implies that for any state $\langle S_x^2\rangle = \langle S_y^2\rangle = \langle S_z^2\rangle = \frac{1}{4}$, which is consistent with the fact that the measurement of any component of $\mathbf{S}$ can produce only $\pm\frac{1}{2}$.

The Stern–Gerlach experiment In 1922, Stern and Gerlach[6] conducted some experiments with silver atoms that most beautifully illustrate the degree to which one can know the orientation of a spin-half object. In addition to this interest, these experiments provide clear examples of the standard procedure for extracting experimental predictions from the formalism of quantum mechanics.

A silver atom is a spin-half object and has a magnetic dipole moment $\boldsymbol{\mu}$. which can be used to track the atom's orientation. In a magnetic field $\mathbf{B}$, a magnetic dipole experiences a force $\nabla(\boldsymbol{\mu}\cdot\mathbf{B})$. Consequently, in a field that varies in strength with position, a dipole that is oriented parallel to $\mathbf{B}$ is drawn in to the region of enhanced $|\mathbf{B}|$, whereas one that is antiparallel to $\mathbf{B}$ is repelled from this region. Stern and Gerlach exploited this effect to

---

[5] Here we are again slightly abusing the notation; S_i are taken to be both the spin operators and their matrix representations.

[6] Gerlach, W. & Stern, O., 1922, *Zeit. f. Physik*, **9**, 349

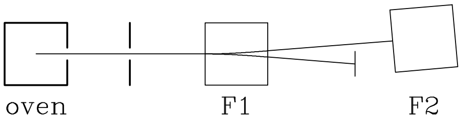


Figure 7.9 Beam split by an SG filter and then up beam hits a second filter.

construct filters along the lines sketched in Figure 7.8. A powerful magnet has one pole sharpened to a knife edge while the other forms either a flat surface (as shown) or is slightly concave. With this geometry the magnetic field lines are close packed as they stream out of the knife edge, and then fan out as they approach the flat pole-piece. Consequently the intensity of the magnetic field increases towards the knife edge and the **Stern–Gerlach filter** sorts particles according to the orientation of their magnetic moments with respect to **B**.

The experiments all start with a beam of sliver atoms moving *in vacuo*, which is produced by allowing vapourised silver to escape from an oven through a suitable arrangement of holes – see Figure 7.9. When the beam passes into a filter, F1, it splits into just two beams of equal intensity. We explain this phenomenon by arguing that the operator μ_i associated with the i^{th} component of an atom's magnetic moment is proportional to S_i: $\mu_i = gS_i$. Hence the filter has 'measured' $\mathbf{n}\cdot\mathbf{S}$, where $\mathbf{n}$ is the unit vector in the direction of $\mathbf{B}$; we are at liberty to orient our coordinate system so that $\mathbf{n} = \mathbf{k}$, and $\mathbf{n}\cdot\mathbf{S} = S_z$. We know that for a spin-half system, a measurement of S_z can yield only $\pm\frac{1}{2}$, so the splitting of the beam into two is explained. Given that there was nothing in the apparatus for producing the beam that favoured up over down as a direction for $\boldsymbol{\mu}$, it is to be expected that half of the atoms return $+\frac{1}{2}$ and half $-\frac{1}{2}$, so the two sub-beams have equal intensity. We block the sub-beam associated with $S_z = -\frac{1}{2}$ so that only particles with $S_z = \frac{1}{2}$ emerge from the filter.

We now place a second Stern-Gerlach filter, F2, in the path of the $|+\rangle$ sub-beam, as shown in Figure 7.9, and investigate the effect of rotating the filter's magnetic axis $\mathbf{n}$ in the plane perpendicular to the incoming beam's direction. Let this be the yz plane. The incoming particles are definitely in the state[7] $|+,z\rangle$ because they've just reported $+\frac{1}{2}$ on a measurement of S_z. F2 measures $\mathbf{n}\cdot\mathbf{S}$, where $\mathbf{n} = (0, \sin\theta, \cos\theta)$ with θ the direction between $\mathbf{n}$ and the z-axis. If $|+,\theta\rangle$ is the eigenket of $\mathbf{n}\cdot\mathbf{S}$ with eigenvalue $+\frac{1}{2}$, the amplitude that the measurement yields $+\frac{1}{2}$ is $\langle +,\theta|+,z\rangle$. The defining equation of $|+,\theta\rangle$ is $\frac{1}{2}\mathbf{n}\cdot\boldsymbol{\sigma}|+,\theta\rangle = \frac{1}{2}|+,\theta\rangle$ or, using the matrix representation (7.110)

$$\begin{pmatrix} \cos\theta & -\mathrm{i}\sin\theta \\ \mathrm{i}\sin\theta & -\cos\theta \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}, \tag{7.113}$$

where $a \equiv \langle +,z|+,\theta\rangle$ and $b \equiv \langle -,z|+,\theta\rangle$. We have to solve this equation subject to the normalisation condition $|a|^2 + |b|^2 = 1$. From the first row of the matrix we deduce that

$$\frac{b}{a} = \mathrm{i}\frac{1-\cos\theta}{\sin\theta}. \tag{7.114}$$

From the trigonometric double-angle formulae we have $1 - \cos\theta = 2\sin^2\frac{1}{2}\theta$ and $\sin\theta = 2\sin\frac{1}{2}\theta\cos\frac{1}{2}\theta$, so

$$\frac{b}{a} = \mathrm{i}\frac{\sin\frac{1}{2}\theta}{\cos\frac{1}{2}\theta.} \tag{7.115}$$

The choices

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \cos\frac{1}{2}\theta \\ \mathrm{i}\sin\frac{1}{2}\theta \end{pmatrix} \tag{7.116}$$

---

[7] We relabel $|+\rangle \to |+,z\rangle$ to make clear that this is a state with spin up along the z-axis.

satisfy both equation (7.115) and the normalisation condition. The amplitude that a particle with spin up along the z-axis also has spin up along the $\mathbf{n}$-axis is $a^* = \langle +,\theta|+,z\rangle$, so the probability that an atom will pass F2 is

$$P_2 = |a|^2 = \cos^2 \tfrac{1}{2}\theta. \tag{7.117}$$

Thus, as θ is increased from 0 to π, the fraction of atoms that get through F2 declines from unity to zero, becoming $\frac{1}{2}$ when $\theta = \pi/2$ and the magnetic axes of F1 and F2 are at right angles. Physically it would be surprising if the fraction that passed F2 when $\theta = \pi/2$ were not a half since, when the magnetic moments of incoming atoms are perpendicular to the magnetic axis of a filter, there is no nothing in the geometry of the experiment to favour the outgoing particles being parallel to the magnetic axis, rather than antiparallel. When $\theta = \pi$ the magnetic axes of the filters are antiparallel and it is obvious that every atom passed by F1 must be blocked by F2. This agrees with what we found out about a spin-half object's orientation in the previous section; if it is pointing somewhere in the upper z hemisphere, then there is some chance it is also pointing in any other hemisphere apart from the $-z$ one.

We now place a third filter, F3, in the atomic beam that emerges from F2. Let ϕ denote the angle between the magnetic axis of this filter and the z-axis. The atoms that emerge from F2 are in the state $|+,\theta\rangle$ because they've just returned $\frac{1}{2}$ on a measurement of $\mathbf{n}\cdot\mathbf{S}$, so the amplitude that these atoms get through F3 is $\langle +,\phi|+,\theta\rangle$. The amplitudes $a' \equiv \langle +,z|+,\phi\rangle$ and $b' \equiv \langle -,z|+,\phi\rangle$ can be obtained directly from the formula we already have for (a,b) with ϕ substituted for θ. Hence

$$\begin{pmatrix} a' \\ b' \end{pmatrix} = \begin{pmatrix} \cos\frac{1}{2}\phi \\ \mathrm{i}\sin\frac{1}{2}\phi \end{pmatrix}. \tag{7.118}$$

and the amplitude to pass F3 is

$$\begin{aligned} \langle +,\phi|+,\theta\rangle &= \sum_{s=\pm} \langle +,\phi|s,z\rangle\langle s,z|+,\theta\rangle \\ &= \left(\cos\tfrac{1}{2}\phi \quad -\mathrm{i}\sin\tfrac{1}{2}\phi\right)\begin{pmatrix} \cos\frac{1}{2}\theta \\ \mathrm{i}\sin\frac{1}{2}\theta \end{pmatrix} \\ &= \cos\tfrac{1}{2}(\phi-\theta). \end{aligned} \tag{7.119}$$

Thus the amplitude to pass F3 depends only on the angle $\phi - \theta$ between the magnetic axes of the filters, and the probability of passing F3 could have been obtained simply by substituting this angle into equation (7.117). This conclusion is obvious physically, but it is satisfying to see it emerge automatically from the formalism.

An especially interesting case is when $\theta = \pi/2$ and $\phi = \pi$. In the absence of F2, F3 would now block every atom that passed F1. But with F2 present both F2 and F3 allow through half of the atoms that reach them, so a quarter of the atoms that leave F1 with $S_z = +\frac{1}{2}$ pass both filters. These atoms exit from F3 with $S_z = -\frac{1}{2}$. Introducing F2 changes the fraction of atoms that pass F3 because the measurement that F2 makes changes the states of the atoms. This is a recurring theme in quantum mechanics. No measurement can be made without slightly disturbing the system that is being measured, and if the system is small enough, the disturbance caused by a measurement can significantly affect the system's dynamics.

7.4.3 Spin-one systems

In the case that $s = 1$, three values of m are possible, $-1, 0, 1$, and so the S_i may be represented by 3×3 matrices. The calculation of these matrices proceeds exactly as for spin half, the main difference being that (7.5) and (7.7) now yield

$$\begin{aligned} S_+|-1\rangle = \sqrt{2}|0\rangle\ ;\ S_+|0\rangle = \sqrt{2}|1\rangle\ ; \\ S_-|1\rangle = \sqrt{2}|0\rangle\ ;\ S_-|0\rangle = \sqrt{2}|-1\rangle\ . \end{aligned} \tag{7.120}$$

The result is

$$\begin{gathered} S_x = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad ; \quad S_y = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix} \\ S_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \end{gathered} \tag{7.121}$$

Consider the effect of using Stern-Gerlach filters on a beam of spin-one atoms. In the experiment depicted in Figure 7.9 each filter now splits the incoming beam into three sub-beams, and we block all but the beam associated with $m = +1$ along the magnetic axis. One third of the atoms that emerge from the collimating slits get through the first filter F1 because each value of m is equally probable at this stage.[8] Therefore one third of the atoms will get through the first filter F1. To calculate the fraction of atoms which then pass through F2, the magnetic axis of which is inclined at angle θ to that of F1, we must calculate the amplitude $\langle 1, \theta | 1, z\rangle$. The defining equation of $|1, \theta\rangle$ is $\mathbf{n} \cdot \mathbf{S}|1, \theta\rangle = |1, \theta\rangle$, which with equations (7.121) can be written

$$\begin{pmatrix} \cos\theta & -\frac{\mathrm{i}}{\sqrt{2}}\sin\theta & 0 \\ \frac{\mathrm{i}}{\sqrt{2}}\sin\theta & 0 & -\frac{\mathrm{i}}{\sqrt{2}}\sin\theta \\ 0 & \frac{\mathrm{i}}{\sqrt{2}}\sin\theta & -\cos\theta \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \tag{7.122}$$

where $a \equiv \langle 1, z|1, \theta\rangle$, $b \equiv \langle 0, z|1, \theta\rangle$, and $c \equiv \langle -1, z|1, \theta\rangle$. The first and third equations, respectively, yield

$$(\cos\theta - 1)a = \frac{\mathrm{i}}{\sqrt{2}}\sin\theta\, b \quad \text{and} \quad \frac{\mathrm{i}}{\sqrt{2}}\sin\theta\, b = (1 + \cos\theta)c. \tag{7.123}$$

Eliminating a and c in favour of b yields

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = b\begin{pmatrix} \frac{\mathrm{i}}{\sqrt{2}}\dfrac{\sin\theta}{\cos\theta - 1} \\ 1 \\ \frac{\mathrm{i}}{\sqrt{2}}\dfrac{\sin\theta}{\cos\theta + 1} \end{pmatrix} = b\begin{pmatrix} -\frac{\mathrm{i}}{\sqrt{2}}\cot(\theta/2) \\ 1 \\ \frac{\mathrm{i}}{\sqrt{2}}\tan(\theta/2) \end{pmatrix}. \tag{7.124}$$

The normalisation condition $|a|^2 + |b|^2 + |c|^2 = 1$ now implies that $b = \sqrt{2}\sin(\frac{1}{2}\theta)\cos(\frac{1}{2}\theta)$. The coefficient that we need is therefore

$$\langle 1, \theta|1, z\rangle = a = \mathrm{i}\cos^2(\tfrac{1}{2}\theta) = \tfrac{\mathrm{i}}{2}(1 + \cos\theta). \tag{7.125}$$

Hence the probability that an atom passes F2 after passing F1 falls from unity when $\theta = 1$ to zero when $\theta = \pi$ as we would expect. When $\theta = \pi/2$ the

---

[8] Here we anticipate a conclusion we'll in §7.5.3: every quantum state has equal a priori probability.

probability is $P_3 = \frac{1}{4}$, which is substantially smaller than the corresponding probability of $\frac{1}{2}$ found in (7.117) for the case of spin-half atoms.

From a classical point of view it is surprising that after F1 has selected atoms that have their angular momentum oriented parallel to the z-axis (in the sense that S_z takes the largest allowed value) there is a non-zero probability P_3 that the angular momentum is subsequently found to be, in the same sense, aligned with the y axis. The explanation of this phenomenon is that for this system, the value of S^2 is $s(s+1) = 2$ which is twice the largest allowed value of S_z. Hence, even in the state $|1, z\rangle$ a significant component of the angular momentum lies in the xy plane. P_3 is the probability that this component is found to be parallel to the y axis. Once the measurement of S_y has been made by F3, the atom is no longer in the state $|1, z\rangle$ and we are no longer certain to obtain 1 if we remeasure S_z.

7.4.4 The classical limit

An electric motor that is, say, 1 cm in diameter and weighs about 10 gm might spin at 100 revolutions per second. Its angular momentum would then be $\sim 10^{-3}\,\mathrm{kg\,m^2\,s^{-1}}$, which is $\sim 10^{31}\hbar$. Thus classical physics works with *extremely* large values of the integers s, m. It is interesting to understand how familiar phenomena emerge from the quantum formalism when s is no longer small.

For any value of s we can construct matrices that represent the angular momentum operators. The matrix for S_z is diagonal with the eigenvalues $s, (s-1), \ldots, -s$ down the diagonal. The matrices for S_x and S_y are evaluated in the usual way from S_+ and S_- and so are zero apart from strips one place above and below the diagonal. Using the relations (7.15) between the coefficients $\alpha_\pm(m)$ of the raising and lowering operators $S_\pm$ we then find

$$\begin{aligned}
S_x &= \frac{1}{2}\begin{pmatrix}
0 & \alpha(s-1) & 0 & \dots & \dots & 0 & 0 \\
\alpha(s-1) & 0 & \alpha(s-2) & & \dots & 0 & 0 \\
0 & \alpha(s-2) & 0 & \ddots & & \vdots & \vdots \\
\vdots & & \ddots & \ddots & \ddots & & \vdots \\
\vdots & \vdots & & \ddots & 0 & \alpha(s-2) & 0 \\
0 & 0 & \dots & & \alpha(s-2) & 0 & \alpha(s-1) \\
0 & 0 & \dots & \dots & 0 & \alpha(s-1) & 0
\end{pmatrix} \\
&= \frac{1}{2}\Big[\alpha(m)\delta_{m,n-1} + \alpha(m-1)\delta_{m,n+1}\Big]
\end{aligned} \tag{7.126a}$$

$$\begin{aligned}
S_y &= \frac{1}{2\mathrm{i}}\begin{pmatrix}
0 & \alpha(s-1) & 0 & \dots & \dots & 0 & 0 \\
-\alpha(s-1) & 0 & \alpha(s-2) & & \dots & 0 & 0 \\
0 & -\alpha(s-2) & 0 & \ddots & & \vdots & \vdots \\
\vdots & & \ddots & \ddots & \ddots & & \vdots \\
\vdots & \vdots & & \ddots & 0 & \alpha(s-2) & 0 \\
0 & 0 & \dots & & -\alpha(s-2) & 0 & \alpha(s-1) \\
0 & 0 & \dots & \dots & 0 & -\alpha(s-1) & 0
\end{pmatrix} \\
&= \frac{1}{2\mathrm{i}}\Big[-\alpha(m)\delta_{m,n-1} + \alpha(m-1)\delta_{m,n+1}\Big]
\end{aligned} \tag{7.126b}$$

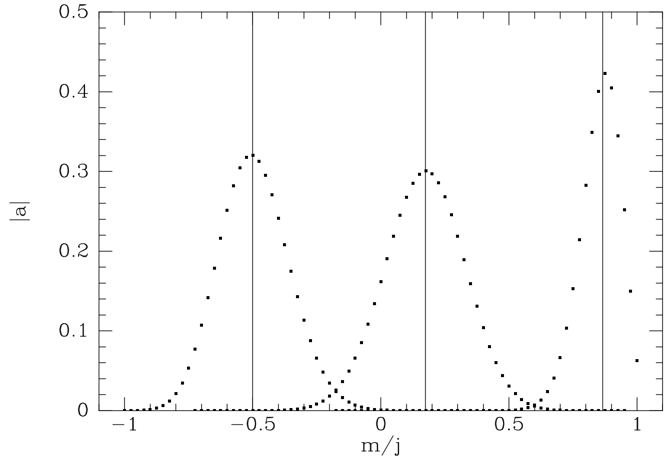


Figure 7.10 The points show the absolute values of the amplitudes $a_m \equiv \langle s,m,z|s,s,\theta\rangle$ for $j=40$ and, from left to right, $\theta=120°, 80°, 30°$. For each value of θ, the vertical line shows the value of $\cos\theta$.

$$
\begin{aligned}
S_z &= \begin{pmatrix}
s & 0 & \dots & 0 & \dots & 0 & 0\\
0 & s-1 & & \vdots & & 0 & 0\\
\vdots & & \ddots & & & \vdots & \vdots\\
0 & & & m & & & 0\\
\vdots & \vdots & & & \ddots & & \vdots\\
0 & 0 & & \vdots & & 1-s & 0\\
0 & 0 & \dots & 0 & \dots & 0 & -s
\end{pmatrix}\\
&= m\ \delta_{mn},
\end{aligned}
\tag{7.126c}
$$

where the $\alpha(m)$ are what were called $\alpha_+(m)$ in (7.15), and the rows and columns of the matrix are labelled from $+s$ at the top left to $-s$ at the bottom right. In the same way as for spins $s=\frac{1}{2}$ and $s=1$, it is straightforward (for a computer) to determine the amplitudes $a_m \equiv \langle s,m,z|s,s,\theta\rangle$ for measuring S_z to have value m, given that $\mathbf{n}\cdot\mathbf{S}$ certainly returns value s when $\mathbf{n}$ is inclined at angle θ to the z-axis. The points in Figure 7.10 show the results of this calculation with $s=40$ and three values of θ. The amplitudes peak around the values of the ordinate, $m=s\cos\theta$, that are marked with vertical lines. The larger the spin, the more sharply the amplitudes are peaked around these lines, so that for the extremely large values of s that are characteristic of macroscopic systems, a_m is significantly non-zero only when m differs negligibly from $s\cos\theta$. Hence, in the classical limit S_z is subject to negligible uncertainty for any value of θ; we are certain to measure the value $s\cos\theta$, which is just the projection of the angular-momentum vector $\langle\mathbf{S}\rangle = s\mathbf{n}$ onto the z-axis.

The classical picture implies that when the angular-momentum vector is tipped in the yz plane at angle θ to the z axis, the value of S_y should be $s\sin\theta$. For this state we evaluate the expectation value $\langle S_y\rangle$ by first multiplying the matrix for S_y on the column vector of the amplitudes plotted in Figure 7.10, and then multiplying the resulting vector by the row vector of the complex conjugates of the amplitudes. The expectation value of S_y in this state is

$$
\begin{aligned}
\langle S_y\rangle &= a_m^*(S_y)_{mn}a_n\\
&= \frac{1}{2\mathrm{i}}\sum_{m,n=-s}^{s} a_m^*\left(\alpha(m)\delta_{m,n-1}+\alpha(m-1)\delta_{m,n+1}\right)a_n\\
&\simeq \frac{1}{2\mathrm{i}}\sum_{m=-s}^{s}\left(\alpha(m)a_m^*a_{m+1}+\alpha(m-1)a_m^*a_{m-1}\right),
\end{aligned}
\tag{7.127}
$$

bearing in mind that $\alpha(s) = \alpha(-s-1) = 0$. For a given value of θ, the amplitudes plotted in Figure 7.10 lie on smooth curves so we can use the approximation $|a_{m-1}| \simeq |a_m| \simeq |a_{m+1}|$. The phases of the a_m increase by $\pi/2$ with successive values of m, so $\langle S_y\rangle$ is real and the two terms (7.127) add. Finally, we exploit the fact that $|a_m|$ is small unless $m \simeq s\cos\theta$ and use the approximation for large s, m that

$$\alpha(m) = \sqrt{s(s+1) - m(m+1)} \simeq \sqrt{s^2 - m^2} \simeq s\sin\theta. \tag{7.128}$$

Combining these approximations with the normalisation condition on the a_m gives

$$\langle S_y\rangle \simeq s\sin\theta \sum_m |a_m|^2 = s\sin\theta \tag{7.129}$$

exactly as classical physics leads us to expect.

To determine the uncertainty in S_y we evaluate the expectation of S_y^2. From equation (7.126b) we find that the matrix S_y^2 has elements

$$\begin{aligned} &-\tfrac14 \sum_{p=-s}^{s} \Big(\alpha(m)\delta_{m,p-1} + \alpha(m-1)\delta_{m,p+1}\Big)\Big(\alpha(p)\delta_{p,n-1} + \alpha(p-1)\delta_{p,n+1}\Big) \\ &\simeq -\tfrac14 \left\{\alpha(m)\alpha(m+1)\delta_{m,n-2} + \alpha(m-1)\alpha(m-3)\delta_{m,n+2}\right. \\ &\qquad\qquad\qquad\qquad \left. - \big(\alpha^2(m) + \alpha^2(m-1)\big)\,\delta_{mn}\right\} \end{aligned} , \tag{7.130}$$

where in going to the second line we have ignored corrections when $m = \pm s$ because the amplitudes for these are negligible anyway. Using the same approximations as before, we now find

$$\begin{aligned} \left\langle S_y^2\right\rangle = \sum_{mn} a_m^*(S_y^2)_{mn}a_n \simeq \sum_m \alpha^2(m)|a_m|^2 \simeq (s\sin\theta)^2 \sum_m |a_m|^2 \\ = s^2\sin^2\theta \simeq \langle S_y\rangle^2 . \end{aligned} \tag{7.131}$$

The uncertainty in S_y, being $\sim \left(\left\langle S_y^2\right\rangle - \langle S_y\rangle^2\right)^{1/2}$ is therefore negligible. A similar calculation shows that both $\langle S_x\rangle$ and $\left\langle S_x^2\right\rangle$ vanish to good accuracy. Thus in the classical limit it is normal for all three components of $\mathbf{S}$ to have small uncertainties. However, it should be noted that S_y can be accurately determined precisely because there is *some* uncertainty in S_z: our calculation on $\langle S_y\rangle$ depends crucially on there being several non-zero amplitudes a_m. Quantum interference between states with different values of S_z is responsible for confining the likely values of S_y to a narrow range.

This is the third time we have found that the familiar world re-emerges through quantum interference between states in which some observable has well-defined values: in §2.3.3 we found that bullets can be assigned positions and momenta simultaneously through interference between states of well-defined momentum, in §3.2 we saw that an excited oscillator moves as a result of quantum interference between states of well-defined energy, and now we find that a gyro has a well defined orientation through quantum interference between states of well-defined angular momentum. In the classical regime a tiny fractional uncertainty in the value of an observable allows the vast numbers of states to have non-negligible amplitudes, and interference between these states narrowly defines the value of the variable that is canonically conjugate to the observable (§2.3.1).

7.5 Addition of angular momenta

In practical applications of quantum mechanics we can often identify two or more components of the system that each carry a well defined amount of angular momentum. For example, in a hydrogen atom both the proton and the electron carry angular momentum $\frac{1}{2}\hbar$ by virtue of their spins, and a further quantity of angular momentum may be present in the orbit of the electron around the proton. The total angular momentum of the atom is the sum of these three contributions, so it is important to understand how to add angular momenta in quantum mechanics. Once we understand how to add two contributions, we'll be able to add any number of contributions, because we can add the third contribution to the result of adding the first two, and so on. Therefore in this section we focus the problem of adding the angular momenta of two 'gyros', that is two systems that have unvarying total angular momentum quantum number j but several possible orientations.

Imagine that we have two gyros in a box and that we know that the first gyro has total angular-momentum quantum number j_1, while the second gyro has total quantum number j_2. Without loss of generality we may assume $j_1 \geq j_2$. A ket describing the state of the first gyro is of the form

$$|\psi_1\rangle = \sum_{m=-j_1}^{j_1} c_m |j_1, m\rangle, \tag{7.132a}$$

while the state of the second is

$$|\psi_2\rangle = \sum_{m=-j_2}^{j_2} d_m |j_2, m\rangle, \tag{7.132b}$$

and from the discussion in §6.1 it follows that the state of the box is

$$|\psi\rangle = |\psi_1\rangle|\psi_2\rangle. \tag{7.133}$$

The coefficients c_m and d_m are the amplitudes to find the individual gyros in particular orientations with respect to the z axis. For example, if both gyros are maximally aligned with the z axis, we will have $|c_{j_1}| = |d_{j_2}| = 1$ and $c_{m_1} = d_{m_2} = 0$ for $m_1 \neq j_1$ and $m_2 \neq j_2$.

The operators of interest are the operators J_i^2, J_{iz} and $J_{i\pm}$ of the i^{th} gyro and the corresponding operators of the box. The operators J_z and $J_\pm$ for the box are simply sums of the corresponding operators for the gyros

$$J_z = J_{1z} + J_{2z} \quad ; \quad J_\pm = J_{1\pm} + J_{2\pm}. \tag{7.134}$$

Operators belonging to different systems always commute, so $[J_{1i}, J_{2j}] = 0$ for any values of i, j. The operator for the square of the box's angular momentum is

$$J^2 = (\mathbf{J}_1 + \mathbf{J}_2)^2 = J_1^2 + J_2^2 + 2\mathbf{J}_1.\mathbf{J}_2\,. \tag{7.135}$$

Now

$$\begin{aligned} J_{1+}J_{2-} &= (J_{1x} + \mathrm{i}J_{1y})(J_{2x} - \mathrm{i}J_{2y}) \\ &= (J_{1x}J_{2x} + J_{1y}J_{2y}) + \mathrm{i}(J_{1y}J_{2x} - J_{1x}J_{2y}). \end{aligned} \tag{7.136}$$

The expression for $J_{1-}J_{2+}$ can be obtained by swapping the labels 1 and 2, so[9]

$$J_{1+}J_{2-} + J_{1-}J_{2+} + 2J_{1z}J_{2z} = 2\mathbf{J}_1.\mathbf{J}_2\,. \tag{7.137}$$

Using this expression to eliminate $\mathbf{J}_1.\mathbf{J}_2$ from (7.135) we obtain

$$J^2 = J_1^2 + J_2^2 + J_{1+}J_{2-} + J_{1-}J_{2+} + 2J_{1z}J_{2z}\,. \tag{7.138}$$

---

[9] Recall that J_{1i} commutes with J_{2j} for all ij.

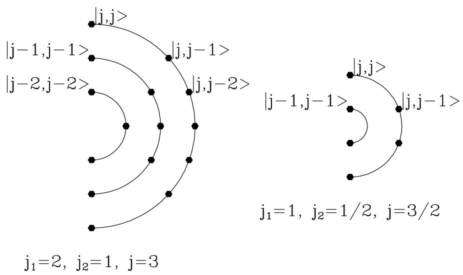


Figure 7.11 The left panel shows states obtained by adding a system of angular momentum $j_2 = 1$ to one with $j_1 = 2$, while the right panel is for $j_1 = 1$ and $j_2 = \frac{1}{2}$.

While the total angular momenta of the individual gyros are fixed, that of the box is variable because it depends on the mutual orientation of the two gyros: if the latter are parallel, the squared angular momentum in the box might be expected to have quantum number $j_1 + j_2$, while if they are antiparallel, the box's angular momentum might be expected to have quantum number $j_1 - j_2$. We shall show that this conjecture is true by explicitly calculating the values of the coefficients c_m and d_m for which the box is in an eigenstate of both J^2 and J_z. We start by examining the state $|j_1, j_1\rangle|j_2, j_2\rangle$ in which both gyros are maximally aligned with the z axis. It is easy to see that this object is an eigenket of J_z with eigenvalue $j_1 + j_2$. We use (7.138) to show that it is also an eigenket of J^2:

$$\begin{aligned} J^2|j_1, j_1\rangle|j_2, j_2\rangle &= (J_1^2 + J_2^2 + J_{1+}J_{2-} + J_{1-}J_{2+} + 2J_{1z}J_{2z})|j_1, j_1\rangle|j_2, j_2\rangle \\ &= \{j_1(j_1+1) + j_2(j_2+1) + 2j_1j_2\}|j_1, j_1\rangle|j_2, j_2\rangle, \end{aligned} \tag{7.139}$$

where we have used the equation $J_{i+}|j_i, j_i\rangle = 0$, which follows from equation (7.7). It is straightforward to show that the expression in curly brackets in equation (7.139) equals $j(j+1)$ with $j = j_1 + j_2$. Hence $|j_1, j_1\rangle|j_2, j_2\rangle$ satisfies both the defining equations of the state $|j_1 + j_2, j_1 + j_2\rangle$ and we may write

$$|j_1 + j_2, j_1 + j_2\rangle = |j_1, j_1\rangle|j_2, j_2\rangle. \tag{7.140}$$

Now that we have found one mutual eigenket for the box of J^2 and J_z we can easily find others by applying J_- to reorient the angular momentum of the box away from the z axis. Again setting $j = j_1 + j_2$ we evaluate the two sides of the equation

$$J_-|j, j\rangle = (J_{1-} + J_{2-})|j_1, j_1\rangle|j_2, j_2\rangle. \tag{7.141}$$

Equation (7.7) enables us to rewrite the left side

$$J_-|j, j\rangle = \sqrt{j(j+1) - j(j-1)}\,|j, j-1\rangle = \sqrt{2j}\,|j, j-1\rangle. \tag{7.142}$$

The right side of (7.141) becomes

$$\begin{aligned} &\sqrt{j_1(j_1+1) - j_1(j_1-1)}|j_1, j_1 - 1\rangle|j_2, j_2\rangle \\ &\quad + \sqrt{j_2(j_2+1) - j_2(j_2-1)}|j_1, j_1\rangle|j_2, j_2 - 1\rangle \\ &\quad = \sqrt{2j_1}\,|j_1, j_1 - 1\rangle|j_2, j_2\rangle + \sqrt{2j_2}\,|j_1, j_1\rangle|j_2, j_2 - 1\rangle. \end{aligned} \tag{7.143}$$

Putting the two sides back together, we have

$$|j, j-1\rangle = \sqrt{\frac{j_1}{j}}\,|j_1, j_1 - 1\rangle|j_2, j_2\rangle + \sqrt{\frac{j_2}{j}}\,|j_1, j_1\rangle|j_2, j_2 - 1\rangle. \tag{7.144}$$

A further application of J_- to the left side of this equation and of $J_{1-} + J_{2-}$ to the right side would produce an expression for $|j, j-2\rangle$ and so on.

Table 7.4 Total numbers of states

j	Number of states
j_1+j_2	$2(j_1+j_2)+1$
j_1+j_2-1	$2(j_1+j_2)+1-2$
...	...
j_1-j_2	$2(j_1+j_2)+1-4j_2$
Total	$(2j_1+1)(2j_2+1)$

Figure 7.11 helps to organise the results of this calculation. States of the box with well defined angular momentum are marked by dots. The radius of each semi-circle is proportional to j', where $j'(j'+1)$ is the eigenvalue of the kets with respect to J^2. The height of each ket above the centre of the circles is proportional to m. The left panel shows the case $j_1=2$, $j_2=1$, while the right panel is for $j_1=1$, $j_2=\frac{1}{2}$. The scheme for constructing eigenstates J^2 and J_z that we have developed so far starts with the state at the top and then uses J_- to successively generate the states that lie on the outermost semi-circle.

We now seek an expression for the state $|j-1,j-1\rangle$ that lies at the top of the first semicircle inwards. It is trivial to verify that $|j_1,m_1\rangle|j_2,m_2\rangle$ is an eigenket of J_z with eigenvalue (m_1+m_2). We require $m_1+m_2=j_1+j_2-1$, so either $m_1=j_1-1$ and $m_2=j_2$, or $m_1=j_1$ and $m_2=j_2-1$. Equation (7.144) shows that $|j,j-1\rangle$ involves precisely these two cases, and must be orthogonal to $|j-1,j-1\rangle$ because it has a different eigenvalue with respect to J^2. So the ket we seek is the unique (up to an overall phase factor) linear combination of the kets appearing in (7.144) that is orthogonal to the linear combination that appears there. That is,

$$|j-1,j-1\rangle=\sqrt{\frac{j_2}{j}}\,|j_1,j_1-1\rangle|j_2,j_2\rangle-\sqrt{\frac{j_1}{j}}\,|j_1,j_1\rangle|j_2,j_2-1\rangle. \qquad (7.145)$$

All the kets $|j-1,m\rangle$ for $m=j-2,\ldots$, which in Figure 7.11 lie on the first semicircle in, can be constructed by applying J_- to this equation.

Similarly, $|j-2,j-2\rangle$, which in Figure 7.11 lies at the top of the second semicircle in, will be a linear combination of $|j_1,j_1-2\rangle|j_2,j_2\rangle$, $|j_1,j_1-1\rangle|j_2,j_2-1\rangle$ and $|j_1,j_1\rangle|j_2,j_2-2\rangle$ and must be orthogonal to $|j,j-2\rangle$ and $|j-1,j-2\rangle$, which are known linear combinations of these states. Hence we can determine which linear combination is required for $|j-2,j-2\rangle$, and then generate the remaining kets of the series $|j-2,m\rangle$ by applying J_- to it.

On physical grounds we would expect the box's smallest total angular momentum quantum number to be j_1-j_2, corresponding to the case in which the two gyros are antiparallel (recall that we have labelled the gyros such that $j_1\geq j_2$). Does this conjectured smallest value of j allow for the correct number of basis states for the box? That is, will there be as many basis states of the box as there are of the contents of the box? We can easily evaluate the latter: there are $2j_1+1$ orientations of the first gyro, and for each of these orientations, the second gyro can be oriented in $2j_2+1$ ways. So the box's contents can be in $(2j_1+1)(2j_2+1)$ basis states of the form $|j_1,m_1\rangle|j_2,m_2\rangle$. The predicted number of basis states of the box is worked out in Table 7.4. In the main part of the table, the number of states in each row is two less than in the row above and there are $2j_2+1$ rows. The sum at the bottom can be obtained by making a third column that is just the second column in reverse order and noting that the sum of the numbers in the second and third columns of a given row is then always $4j_1+2$. Hence twice the sum of the numbers in the second column is $2j_2+1$ times $4j_1+2$. Thus we *do* get the correct number of basis states if the smallest value of j is j_1-j_2.

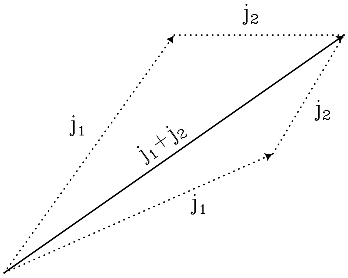


Figure 7.12 Interpretation of Clebsch–Gordan coefficients in terms of vectors. The full line has length $\sqrt{3(3+1)}$ and its vertical component has length 2. The dotted lines labelled j_1 have length $\sqrt{2(2+1)}$ and vertical components of length 2 and 1.

The numbers

$$C(j,m;j_1,j_2,m_1,m_2) \equiv \langle j,m|j_1,m_1\rangle|j_2,m_2\rangle \tag{7.146}$$

that we have been evaluating are called **Clebsch–Gordan** coefficients. They have a simple physical interpretation: $C(j,m;j_1,j_2,m_1,m_2)$ is the amplitude that, on opening the box when it's in a state of well defined angular momentum, we will find the first and second gyros to be oriented with amounts m_1 and m_2 of their spins parallel to the z axis. For example, equation (7.145) implies that $C(3,2;2,1,1,1) = \sqrt{2/3}$, so if a box that contains a spin-two gyro and a spin-one gyro has spin-three, there is a probability 2/3 that on opening the box the second gyro will be maximally aligned with the z axis and the second significantly inclined, and only a probability 1/3 of finding the reverse arrangement. These two possibilities are depicted by the lower and upper dotted lines in Figure 7.12. The classical interpretation is that the two gyros precess around the fixed angular-momentum vector of the box, and that the two configurations for which the Clebsch–Gordan coefficients give amplitudes are two of the states through which the precession carries the system. This picture is intuitive and of some value, but should not be taken too seriously. For one thing, the rules for adding angular momentum are independent of any statement about the Hamiltonian, and therefore carry no implication about the time evolution of the system. The gyros may or may not precess, depending on whether they are dynamically coupled.

In §6.1 we saw that the physical significance of the state of a composite system, such as that formed by two gyros, being a linear combination of product states such as $|j_1,m_1\rangle|j_2,m_2\rangle$ is that the subsystems are correlated. The Clebsch–Gordan coefficients encode the correlations between the gyros required for the box to have well-defined angular momentum. If there is any uncertainty in the orientation of either gyro, such correlations are essential if the angular momentum of the box is to be well defined: the angular momentum of the second gyro has to make up a pre-defined total with whatever value is measured for the first gyro. This consideration explains why the only states of the box that are simple products of states of the individual gyros are $|j_1+j_2,j_1+j_2\rangle = |j_1,j_1\rangle|j_2,j_2\rangle$ and $|j_1+j_2,-(j_1+j_2)\rangle = |j_1,-j_1\rangle|j_2,-j_2\rangle$ – so much angular momentum can be aligned with the z-axis only by each gyro individually straining to align with the axis, and there is then no need for the gyros to coordinate their efforts.

7.5.1 Case of two spin-half systems

The general analysis we have just given will be clarified by working out some particular cases. We consider first the case $j_1 = j_2 = \frac{1}{2}$, which is relevant, for example, to a hydrogen atom in its ground state, when all angular momentum is contributed by the spins of the proton and the electron. The electron has base states $|\pm,\mathrm{e}\rangle$ in which J_z returns the value $\pm\frac{1}{2}$, while the proton has corresponding base states $|\pm,\mathrm{p}\rangle$. Hence there are four states in all and j takes just two values, 1 and 0.

Our construction of the states in which the atom has well-defined angular momentum starts with the state

$$|1,1\rangle = |+,\mathrm{e}\rangle|+,\mathrm{p}\rangle \tag{7.147}$$

in which both the electron and the proton have their spins maximally aligned with the z axis. So the atom has maximum angular momentum, and its angular momentum is maximally aligned with the z axis. Applying $J_- = J_-^{\mathrm{e}} + J_-^{\mathrm{p}}$ to this ket we obtain

$$|1,0\rangle = \frac{1}{\sqrt{2}}\left(|-,\mathrm{e}\rangle|+,\mathrm{p}\rangle + |+,\mathrm{e}\rangle|-,\mathrm{p}\rangle\right). \tag{7.148}$$

The right side of this equation states that with the atom in this state, measurements of J_z for the electron and proton are certain to find that they are 'antiparallel'. This fact is surprising given that the left side states that the atom has maximum angular momentum, so you would think that the two particles had parallel angular momenta. The resolution of this paradox is that the z components of the two spins are antiparallel, but the components in the xy plane are parallel, although their direction is unknown to us. Similarly, when the atom is in the state $|1,1\rangle$ of equation (7.147), the z components of the electron and proton angular momenta are parallel, but the components in the xy plane are not well aligned. The poor alignment in the xy plane explains why $\sqrt{J^2} = \sqrt{2}$ for the atom is less than $\sqrt{3}$, which is the sum of $\sqrt{J^2} = \sqrt{3/4}$ for the electron and the proton.

When we apply J_- to $|1,0\rangle$ we obtain

$$|1,-1\rangle = |-,\mathrm{e}\rangle|-,\mathrm{p}\rangle. \tag{7.149}$$

This equation confirms the physically obvious fact that if we want to have $\hbar$ of angular momentum pointing along the negative z axis, we need to have the angular momenta of both the proton and the electron maximally aligned with the negative z axis.

The remaining state of the atom is $|0,0\rangle$ in which the atom has no angular momentum. This is the unique linear combination of the two compound states on the right of equation (7.148) that is orthogonal to $|1,0\rangle$:

$$|0,0\rangle = \frac{1}{\sqrt{2}}\left(|-,\mathrm{e}\rangle|+,\mathrm{p}\rangle - |+,\mathrm{e}\rangle|-,\mathrm{p}\rangle\right). \tag{7.150}$$

The change of sign on the right of this equation from the right of equation (7.148) for $|1,0\rangle$ ensures that the spins of the electron and proton are antiparallel in the xy plane as well as along the z axis. We show this by rewriting $|1,0\rangle$ and $|0,0\rangle$ in terms of the states in which the electron and proton have well-defined spin parallel to the x-axis. These states are

$$\begin{aligned}
|x+,\mathrm{e}\rangle &= \frac{1}{\sqrt{2}}\left(|+,\mathrm{e}\rangle + |-,\mathrm{e}\rangle\right) \qquad & |x-,\mathrm{e}\rangle &= \frac{1}{\sqrt{2}}\left(|+,\mathrm{e}\rangle - |-,\mathrm{e}\rangle\right)\\
|x+,\mathrm{p}\rangle &= \frac{1}{\sqrt{2}}\left(|+,\mathrm{p}\rangle + |-,\mathrm{p}\rangle\right) \qquad & |x-,\mathrm{p}\rangle &= \frac{1}{\sqrt{2}}\left(|+,\mathrm{p}\rangle - |-,\mathrm{p}\rangle\right)
\end{aligned} \tag{7.151}$$

So

$$\begin{aligned}
|0,0\rangle &= |x+,\mathrm{e}\rangle\langle x+,\mathrm{e}|0,0\rangle + |x-,\mathrm{e}\rangle\langle x-,\mathrm{e}|0,0\rangle\\
&= \tfrac{1}{2}|x+,\mathrm{e}\rangle\left(-|-,\mathrm{p}\rangle + |+,\mathrm{p}\rangle\right) - \tfrac{1}{2}|x-,\mathrm{e}\rangle\left(|-,\mathrm{p}\rangle + |+,\mathrm{p}\rangle\right)\\
&= -\frac{1}{\sqrt{2}}\left(|x+,\mathrm{e}\rangle|x-,\mathrm{p}\rangle + |x-,\mathrm{e}\rangle|x+,\mathrm{p}\rangle\right).
\end{aligned} \tag{7.152}$$

The last line states that when the atom is in the state $|0,0\rangle$ we are indeed guaranteed to find the components of the spins of the electron and proton parallel to x have opposite signs. An analogous calculation starting from equation (7.148) yields (Problem 7.16)

$$|1,0\rangle = \frac{1}{\sqrt{2}}\left(|x+,\mathrm{e}\rangle|x+,\mathrm{p}\rangle - |x-,\mathrm{e}\rangle|x-,\mathrm{p}\rangle\right), \tag{7.153}$$

so when the atom is in the $|1,0\rangle$ state the two particles have identical components of spin along x .

Notice that all three states in which the atom has $j=1$ are unchanged if we swap the m values of the particles – that is, if we map $|\pm,\mathrm{e}\rangle \to |\mp,\mathrm{e}\rangle$ and the same for the proton states. The atomic atomic state with $j=0$, by contrast, changes sign under this interchange. This fact will prove to be important when we consider systems with two electrons (such a helium) or two protons (such as an H_2 molecule).

7.5.2 Case of spin one and spin half

In the first excited state of hydrogen, the electron can have total orbital angular momentum quantum number $l=1$. So we now consider how to combine angular momentum $j=1$ with the electron's spin, $j=\frac{1}{2}$. The total angular momentum quantum number takes two values, $j=\frac{3}{2}$ and $j=\frac{1}{2}$ (see Figure 7.11). We start with the state

$$|\tfrac{3}{2},\tfrac{3}{2}\rangle = |+\rangle|1,1\rangle \tag{7.154}$$

in which the spin and orbital angular momenta are both maximally oriented along the z axis. Applying $J_- = L_- + S_-$ to this equation, we obtain

$$|\tfrac{3}{2},\tfrac{1}{2}\rangle = \sqrt{\tfrac{1}{3}}\,|-\rangle|1,1\rangle + \sqrt{\tfrac{2}{3}}\,|+\rangle|1,0\rangle. \tag{7.155}$$

The right side of this equation says that in this state of the atom, the electron is twice as likely to be found with its spin up as down. A second application of J_- yields

$$|\tfrac{3}{2},-\tfrac{1}{2}\rangle = \sqrt{\tfrac{2}{3}}\,|-\rangle|1,0\rangle + \sqrt{\tfrac{1}{3}}\,|+\rangle|1,-1\rangle \tag{7.156}$$

as we would expect from the symmetry between up and down. A final application of J_- yields $|\frac{3}{2},-\frac{3}{2}\rangle = |-\rangle|1,-1\rangle$ as it must on physical grounds.

The state $|\frac{1}{2},\frac{1}{2}\rangle$ is the linear combination of the states that appear in the right of equation (7.155) that is orthogonal to $|\frac{3}{2},\frac{1}{2}\rangle$. Hence,

$$|\tfrac{1}{2},\tfrac{1}{2}\rangle = \sqrt{\tfrac{2}{3}}\,|-\rangle|1,1\rangle - \sqrt{\tfrac{1}{3}}\,|+\rangle|1,0\rangle. \tag{7.157}$$

In this atomic state, the electron's spin is twice as likely to be down as up. The last remaining state can be found by applying J_- to equation (7.157). It is

$$|\tfrac{1}{2},-\tfrac{1}{2}\rangle = \sqrt{\tfrac{1}{3}}\,|-\rangle|1,0\rangle - \sqrt{\tfrac{2}{3}}\,|+\rangle|1,-1\rangle. \tag{7.158}$$

7.5.3 The classical limit

In classical physics we identify angular momentum with the vector $\boldsymbol{\mathcal{J}} \equiv \hbar\,\langle \mathbf{J}\rangle$, and the angular momentum of the whole system is obtained by vectorially adding the angular momenta $\boldsymbol{\mathcal{J}}_1$ and $\boldsymbol{\mathcal{J}}_2$ of the component parts. If θ is the angle between these vectors, then

$$\mathcal{J}^2 = \mathcal{J}_1^2 + \mathcal{J}_2^2 + 2\mathcal{J}_1\mathcal{J}_2\cos\theta. \tag{7.159}$$

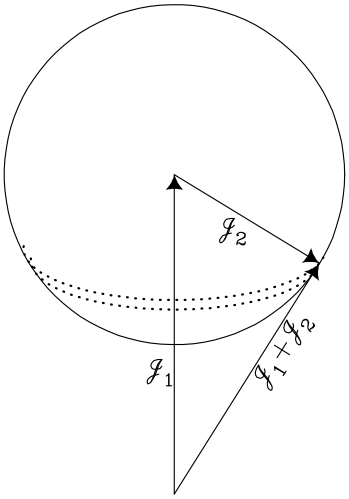


Figure 7.13 Classically the sum vector $\boldsymbol{\mathcal{J}}_1 + \boldsymbol{\mathcal{J}}_2$ can line anywhere on the sphere of radius $|\boldsymbol{\mathcal{J}}_2|$ around the end of $\boldsymbol{\mathcal{J}}_1$.

If nothing is known about the direction of $\boldsymbol{\mathcal{J}}_2$ relative to $\boldsymbol{\mathcal{J}}_1$, all points on a sphere of radius $\mathcal{J}_2$ and centred on the end of $\boldsymbol{\mathcal{J}}_1$ are equally likely locations for the end of $\boldsymbol{\mathcal{J}}_2$ (Figure 7.13). Consequently, the probability $\mathrm{d}P$ that θ lies in the range $(\theta, \theta + \mathrm{d}\theta)$ is proportional to the area of the band shown in the figure. Quantitatively

$$\mathrm{d}P = \tfrac{1}{2}\sin\theta\,\mathrm{d}\theta, \tag{7.160}$$

where the factor $\frac{1}{2}$ ensures that $\int \mathrm{d}P = 1$. From equation (7.159) the change in $\mathcal{J}$ when θ changes by $\mathrm{d}\theta$ is given by

$$\mathcal{J}\,\mathrm{d}\mathcal{J} = -\mathcal{J}_1\mathcal{J}_2\sin\theta\mathrm{d}\theta. \tag{7.161}$$

Combining equations (7.160) and (7.161), we find that the probability that the total angular momentum lies in the interval $(\mathcal{J}, \mathcal{J} + \mathrm{d}\mathcal{J})$ is[10]

$$\mathrm{d}P = \frac{\mathcal{J}\,\mathrm{d}\mathcal{J}}{2\mathcal{J}_1\mathcal{J}_2}. \tag{7.162}$$

In quantum mechanics the fraction of states that have total angular-momentum quantum number j is

$$f = \frac{2j+1}{(2j_1+1)(2j_2+1)}, \tag{7.163}$$

which in the classical limit of large quantum numbers becomes approximately $j/(2j_1j_2)$. If all states were equally likely, this fraction would equal the classical probability that $\mathcal{J} \simeq j\hbar$ lay within $\hbar$ of $j\hbar$. It is easy to check from (7.161) that $\mathrm{d}P$ does indeed take the value f when we insert $\mathcal{J}_i = \hbar j_i$ and $\mathrm{d}\mathcal{J} = \hbar$. Thus from consistency with classical mechanics we are led to the **principle of equal a priori probability**, namely that *when we have no information relevant to an upcoming measurement, we assign equal probabilities to the system being in each state of whatever basis we have decided to work in.* This principle is the foundation of all statistical physics.

---

[10] We discarded the minus sign in equation (7.161) because we require $\mathrm{d}P > 0$ regardless of whether $\mathcal{J}$ increases or decreases.

Problems

7.1 Show that L_i commutes with $\mathbf{x}\cdot\mathbf{p}$ and thus also with scalar functions of $\mathbf{x}$ and $\mathbf{p}$.

7.2 In the rotation spectrum of $^{12}\mathrm{C}^{16}\mathrm{O}$ the line arising from the transition $l = 4 \to 3$ is at 461.04077 GHz, while that arising from $l = 36 \to 35$ is at 4115.6055 GHz. Show from these data that in a non-rotating CO molecule the intra-nuclear distance is $s \simeq 0.113\,\mathrm{nm}$, and that the electrons provide a spring between the nuclei that has force constant $\sim 1904\,\mathrm{N\,m^{-1}}$. Hence show that the vibrational frequency of CO should lie near $6.47 \times 10^{13}\,\mathrm{Hz}$ (measured value is $6.43 \times 10^{13}\,\mathrm{Hz}$). Hint: show from classical mechanics that the distance of O from the centre of mass is $\frac{3}{7}s$ and that the molecule's moment of inertia is $\frac{48}{7}m_{\mathrm{p}}s^2$. Recall also the classical relation $L = I\omega$.

7.3* We have that

$$L_+ \equiv L_x + \mathrm{i}L_y = \mathrm{e}^{\mathrm{i}\phi}\Big(\frac{\partial}{\partial\theta} + \mathrm{i}\cot\theta\frac{\partial}{\partial\phi}\Big). \tag{7.164}$$

From the Hermitian nature of $L_z = -\mathrm{i}\partial/\partial\phi$ we infer that derivative operators are anti-Hermitian. So using the rule $(AB)^\dagger = B^\dagger A^\dagger$ on equation (7.164), we infer that

$$L_- \equiv L_+^\dagger = \Big(-\frac{\partial}{\partial\theta} + \mathrm{i}\frac{\partial}{\partial\phi}\cot\theta\Big)\mathrm{e}^{-\mathrm{i}\phi}.$$

This argument and the result it leads to is wrong. Obtain the correct result by integrating by parts $\int \mathrm{d}\theta\,\sin\theta \int \mathrm{d}\phi\,(f^* L_+ g)$, where f and g are arbitrary functions of θ and ϕ. What is the fallacy in the given argument?

7.4* By writing $\hbar^2L^2 = (\mathbf{x}\times\mathbf{p})\cdot(\mathbf{x}\times\mathbf{p}) = \sum_{ijklm} x_j p_k\,\epsilon_{ilm} x_l p_m$ show that

$$p^2 = \frac{\hbar^2L^2}{r^2} + \frac{1}{r^2}\left\{(\mathbf{r}\cdot\mathbf{p})^2 - \mathrm{i}\hbar\mathbf{r}\cdot\mathbf{p}\right\}. \tag{7.165}$$

By showing that $\mathbf{p}\cdot\hat{\mathbf{r}} - \hat{\mathbf{r}}\cdot\mathbf{p} = -2\mathrm{i}\hbar/r$, obtain $\mathbf{r}\cdot\mathbf{p} = rp_r + \mathrm{i}\hbar$. Hence obtain

$$p^2 = p_r^2 + \frac{\hbar^2L^2}{r^2}. \tag{7.166}$$

Give a physical interpretation of one over $2m$ times this equation.

7.5 The angular part of a system's wavefunction is

$$\langle\theta,\phi|\psi\rangle \propto (\sqrt{2}\cos\theta + \sin\theta\mathrm{e}^{-\mathrm{i}\phi} - \sin\theta\mathrm{e}^{\mathrm{i}\phi}).$$

What are the possible results of measurement of (a) L^2, and (b) L_z, and their probabilities? What is the expectation value of L_z?

7.6 A system's wavefunction is proportional to $\sin^2\theta\,\mathrm{e}^{2\mathrm{i}\phi}$. What are the possible results of measurements of (a) L_z and (b) L^2?

7.7 A system's wavefunction is proportional to $\sin^2\theta$. What are the possible results of measurements of (a) L_z and (b) L^2? Give the probabilities of each possible outcome.

7.8 A system that has total orbital angular momentum $\sqrt{6}\hbar$ is rotated through an angle ϕ around the z axis. Write down the 5×5 matrix that updates the amplitudes a_m that L_z will take the value m.

7.9 Consider a stationary state $|E,l\rangle$ of a free particle of mass m that has angular-momentum quantum number l. Show that $H_l|E,l\rangle = E|E,l\rangle$, where

$$H_l \equiv \frac{1}{2m}\left(p_r^2 + \frac{l(l+1)\hbar^2}{r^2}\right). \tag{7.167}$$

Give a physical interpretation of the two terms in the big bracket. Show that $H_l = A_l^\dagger A_l$, where

$$A_l \equiv \frac{1}{\sqrt{2m}}\left(\mathrm{i}p_r - \frac{(l+1)\hbar}{r}\right). \tag{7.168}$$

Show that $[A_l, A_l^\dagger] = H_{l+1} - H_l$. What is the state $A_l|E,l\rangle$? Show that for $E > 0$ there is no upper bound on the angular momentum. Interpret this result physically.

7.10 Write down the expression for the commutator $[\sigma_i, \sigma_j]$ of two Pauli matrices. Show that the anticommutator of two Pauli matrices is

$$\{\sigma_i, \sigma_j\} = 2\delta_{ij}. \tag{7.169}$$

7.11 Let $\mathbf{n}$ be any unit vector and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ be the vector whose components are the Pauli matrices. Why is it physically necessary that $\mathbf{n}\cdot\boldsymbol{\sigma}$ satisfy $(\mathbf{n}\cdot\boldsymbol{\sigma})^2 = I$, where I is the 2×2 identity matrix? Let $\mathbf{m}$ be a unit vector such that $\mathbf{m}\cdot\mathbf{n} = 0$. Why do we require that the commutator $[\mathbf{m}\cdot\boldsymbol{\sigma}, \mathbf{n}\cdot\boldsymbol{\sigma}] = 2\mathrm{i}(\mathbf{m}\times\mathbf{n})\cdot\boldsymbol{\sigma}$? Prove that that these relations follow from the algebraic properties of the Pauli matrices. You should be able to show that $[\mathbf{m}\cdot\boldsymbol{\sigma}, \mathbf{n}\cdot\boldsymbol{\sigma}] = 2\mathrm{i}(\mathbf{m}\times\mathbf{n})\cdot\boldsymbol{\sigma}$ for any two vectors $\mathbf{n}$ and $\mathbf{m}$.

7.12 Let $\mathbf{n}$ be the unit vector in the direction with polar coordinates (θ, ϕ). Write down the matrix $\mathbf{n}\cdot\boldsymbol{\sigma}$ and find its eigenvectors. Hence show that the state of a spin-half particle in which a measurement of the component of spin along $\mathbf{n}$ is certain to yield $\frac{1}{2}\hbar$ is

$$|+,\mathbf{n}\rangle = \sin(\theta/2)\,\mathrm{e}^{\mathrm{i}\phi/2}|-\rangle + \cos(\theta/2)\,\mathrm{e}^{-\mathrm{i}\phi/2}|+\rangle, \tag{7.170}$$

where $|\pm\rangle$ are the states in which $\pm\frac{1}{2}$ is obtained when s_z is measured. Obtain the corresponding expression for $|-,\mathbf{n}\rangle$. Explain physically why the amplitudes in (7.170) have modulus $2^{-1/2}$ when $\theta = \pi/2$ and why one of the amplitudes vanishes when $\theta = \pi$.

7.13 For a spin-half particle at rest, the rotation operator $\mathbf{J}$ is equal to the spin operator $\mathbf{S}$. Use the result of Problem 7.10 to show that in this case the rotation operator $U(\boldsymbol{\alpha}) \equiv \exp(-\mathrm{i}\boldsymbol{\alpha}\cdot\mathbf{J})$ is

$$U(\boldsymbol{\alpha}) = I\cos\left(\frac{\alpha}{2}\right) - \mathrm{i}\hat{\boldsymbol{\alpha}}\cdot\boldsymbol{\sigma}\sin\left(\frac{\alpha}{2}\right), \tag{7.171}$$

where $\hat{\boldsymbol{\alpha}}$ is the unit vector parallel to $\boldsymbol{\alpha}$. Comment on the value this gives for $U(\boldsymbol{\alpha})$ when $\alpha = 2\pi$.

7.14 Write down the 3×3 matrix that represents S_x for a spin-one system in the basis in which S_z is diagonal (i.e., the basis states are $|0\rangle$ and $|\pm\rangle$ with $S_z|+\rangle = |+\rangle$, etc.)

A beam of spin-one particles emerges from an oven and enters a Stern–Gerlach filter that passes only particles with $J_z = \hbar$. On exiting this filter, the beam enters a second filter that passes only particles with $J_x = \hbar$, and then finally it encounters a filter that passes only particles with $J_z = -\hbar$. What fraction of the particles stagger right through?

7.15 A box containing two spin-one gyros A and B is found to have angular-momentum quantum numbers $j = 2$, $m = 1$. Determine the probabilities that when J_z is measured for gyro A, the values $m = \pm 1$ and 0 will be obtained.

What is the value of the Clebsch–Gordan coefficient $C(2,1;1,1,1,0)$?

7.16 The angular momentum of a hydrogen atom in its ground state is entirely due to the spins of the electron and proton. The atom is in the state $|1,0\rangle$ in which it has one unit of angular momentum but none of it is parallel to the z-axis. Express this state as a linear combination of products of the spin states $|\pm,\mathrm{e}\rangle$ and $|\pm,\mathrm{p}\rangle$ of the proton and electron. Show that the states $|x\pm,\mathrm{e}\rangle$ in which the electron has well-defined spin along the x-axis are

$$|x\pm,\mathrm{e}\rangle = \frac{1}{\sqrt{2}}\left(|+,\mathrm{e}\rangle \pm |-,\mathrm{e}\rangle\right). \tag{7.172}$$

By writing

$$|1,0\rangle = |x+,\mathrm{e}\rangle\langle x+,\mathrm{e}|1,0\rangle + |x-,\mathrm{e}\rangle\langle x-,\mathrm{e}|1,0\rangle, \tag{7.173}$$

express $|1,0\rangle$ as a linear combination of the products $|x\pm,\mathrm{e}\rangle|x\pm,\mathrm{p}\rangle$. Explain the physical significance of your result.

7.17* Repeat the analysis of Problem 7.14 for spin-one particles coming on filters aligned successively along $+z$, 45° from z towards x [i.e. along (1,0,1)], and along x.

Use classical electromagnetic theory to determine the outcome in the case that the spin-one particles were photons and the filters were polaroid. Why do you get a different answer? Which answer is correct?

7.18* Show that l excitations can be divided amongst the x, y or z oscillators of a three-dimensional harmonic oscillator in $(\frac{1}{2}l+1)(l+1)$ ways. Verify in the case $l=4$ that this agrees with the number of states of well defined angular momentum and the given energy.

7.19* Let

$$A_l \equiv \frac{1}{\sqrt{2m\hbar\omega}}\left(\mathrm{i}p_r - \frac{(l+1)\hbar}{r} + m\omega r\right). \tag{7.174}$$

be the ladder operator of the three-dimensional harmonic oscillator and $|E,l\rangle$ be the oscillator's stationary state of energy E and angular-momentum quantum number l. Show that if we write $A_l|E,l\rangle = \alpha_-|E-\hbar\omega, l+1\rangle$, then $\alpha_- = \sqrt{\mathcal{L}-l}$, where $\mathcal{L}$ is the angular-momentum quantum number of a circular orbit of energy E. Show similarly that if $A^\dagger_{l-1}|E,l\rangle = \alpha_+|E+\hbar\omega, l-1\rangle$, then $\alpha_+ = \sqrt{\mathcal{L}-l+2}$.

7.20* Show that the probability distribution in radius of a particle that orbits in the three-dimensional harmonic-oscillator potential on a circular orbit with angular-momentum quantum number l peaks at $r/a = \sqrt{l+1}$, where

$$a \equiv \sqrt{\frac{\hbar}{m\omega}}. \tag{7.175}$$

Derive the corresponding classical result.

7.21* A particle moves in the three-dimensional harmonic oscillator potential with the second largest angular-momentum quantum number possible at its energy. Show that the radial wavefunction is

$$u_1 \propto x^l\left(x - \frac{l+\frac{1}{2}}{x}\right)\mathrm{e}^{-x^2/2} \quad \text{where} \quad x \equiv r/a \quad \text{with} \quad a \equiv \sqrt{\frac{\hbar}{m\omega}}. \tag{7.176}$$

How many radial nodes does this wavefunction have?

7.22* The interaction between neighbouring spin-half atoms in a crystal is described by the Hamiltonian

$$H = K\left(\frac{\mathbf{S}^{(1)}\cdot\mathbf{S}^{(2)}}{a} - 3\frac{(\mathbf{S}^{(1)}\cdot\mathbf{a})(\mathbf{S}^{(2)}\cdot\mathbf{a})}{a^3}\right), \tag{7.177}$$

where K is a constant, $\mathbf{a}$ is the separation of the atoms and $\mathbf{S}^{(1)}$ is the first atom's spin operator. Explain what physical idea underlies this form of H. Show that $S_x^{(1)}S_x^{(2)}+S_y^{(1)}S_y^{(2)}=\frac{1}{2}(S_+^{(1)}S_-^{(2)}+S_-^{(1)}S_+^{(2)})$. Show that the mutual eigenkets of the total spin operators S^2 and S_z are also eigenstates of H and find the corresponding eigenvalues.

At time $t=0$ particle 1 has its spin parallel to $\mathbf{a}$, while the other particle's spin is antiparallel to $\mathbf{a}$. Find the time required for both spins to reverse their orientations.

7.23* Show that $[J_i, L_j] = \mathrm{i}\sum_k \epsilon_{ijk}L_k$ and $[J_i, L^2] = 0$ by eliminating L_i using its definition $\mathbf{L} = \hbar^{-1}\mathbf{x}\times\mathbf{p}$, and then using the commutators of J_i with $\mathbf{x}$ and $\mathbf{p}$.

7.24* In this problem you show that many matrix elements of the position operator $\mathbf{x}$ vanish when states of well defined l, m are used as basis states. These results will lead to selection rules for electric dipole radiation. First show that $[L^2, x_i] = \mathrm{i}\sum_{jk}\epsilon_{jik}(L_jx_k + x_kL_j)$. Then show that $\mathbf{L}\cdot\mathbf{x} = 0$ and using this result derive

$$[L^2,[L^2,x_i]] = \mathrm{i}\sum_{jk}\epsilon_{jik}\big(L_j[L^2,x_k]+[L^2,x_k]L_j\big) = 2(L^2x_i + x_iL^2). \tag{7.178}$$

By squeezing this equation between angular-momentum eigenstates $\langle l, m|$ and $|l', m'\rangle$ show that

$$0 = \big\{(\beta-\beta')^2 - 2(\beta+\beta')\big\}\langle l,m|x_i|l',m'\rangle,$$

where $\beta \equiv l(l+1)$ and $\beta' = l'(l'+1)$. By equating the factor in front of $\langle l,m|x_i|l',m'\rangle$ to zero, and treating the resulting equation as a quadratic equation for β given β', show that $\langle l,m|x_i|l',m'\rangle$ must vanish unless $l+l' = 0$ or $l = l'\pm 1$. Explain why the matrix element must also vanish when $l = l' = 0$.

8

Hydrogen

Wherever we look, down at ourselves or up into the vastness of the Universe, what we see are atoms. The way atoms interact with themselves and with electromagnetic radiation structures the world about us, giving colour, texture, solidity or fluidity to all things, both alive and inanimate. In the wider Universe the way visible matter has aggregated into stars and galaxies is determined by the interplay between atoms and radiation. In the last two decades of the twentieth century it emerged that atoms do not really dominate the Universe; on large scales they are rather like icing on the cake. But they certainly dominate planet Earth, and, like the icing, they are all we can see of the cosmic cake.

Besides the inherent interest of atomic structure, there is the historical fact that the formative years of quantum mechanics were dominated by experimental investigations of atomic structure. Most of the principles of the subject were developed to explain atomic phenomena, and the stature of these phenomena in the minds of physicists was greatly enhanced through the rôle they played in revolutionising physics.

It is an unfortunate fact that atoms are complex systems that are not easily modelled to a precision as good as that with which they are commonly measured. The complexity of an atom increases with the number of electrons that it contains, both because the electrons interact with one another as well as with the nucleus, and because the more electrons there are, the higher the nuclear charge and the faster electrons can move. By the middle of the periodic table the speeds of the fastest electrons are approaching the speed of light and relativistic effects are important.

In this chapter we develop a model of the simplest atom, hydrogen, that accounts for most, but not all, measurements. In Chapter 10 we will take the first steps towards a model of the second most complex atom, helium, and indicate general trends in atomic properties as one proceeds down the periodic table. The ideas we use will depend heavily on the model of hydrogen-like systems that is developed in this chapter. With these applications in view, we generalise from hydrogen to a **hydrogen-like ion**, in which a single electron is bound to a nucleus of charge Ze.

8.1 Gross structure of hydrogen

We start with a rather crude model of a hydrogen-like ion. In this model neither the electron nor the nucleus has a spin, and the electron moves non-relativistically under purely electrostatic forces. The structure of an atom or ion that is obtained using these approximations is called its **gross structure**. The approximations make it easy to write down the model Hamiltonian because they include just three contributions to the energy: the kinetic energies of the nucleus and the electron, and the bodies' electrostatic binding energy:

$$H = \frac{p_{\mathrm{n}}^2}{2m_{\mathrm{n}}} + \frac{p_{\mathrm{e}}^2}{2m_{\mathrm{e}}} - \frac{Ze^2}{4\pi\epsilon_0|\mathbf{x}_{\mathrm{e}} - \mathbf{x}_{\mathrm{n}}|}, \tag{8.1}$$

where $\mathbf{x}_{\mathrm{e}}$ and $\mathbf{x}_{\mathrm{n}}$ are the position operators of the electron and the nucleus, respectively, and $\mathbf{p}_{\mathrm{e}}$ and $\mathbf{p}_{\mathrm{n}}$ are the corresponding momentum operators. We wish to solve the eigenvalue equation $H|E\rangle = E|E\rangle$ for this Hamiltonian. In the position representation, the momentum operators become derivative operators, and the eigenvalue equation becomes a partial differential equation in six variables

$$E\psi(\mathbf{x}_{\mathrm{n}}, \mathbf{x}_{\mathrm{e}}) = -\frac{\hbar^2}{2m_{\mathrm{n}}}\nabla_{\mathrm{n}}^2\psi - \frac{\hbar^2}{2m_{\mathrm{e}}}\nabla_{\mathrm{e}}^2\psi - \frac{Ze^2\psi}{4\pi\epsilon_0|\mathbf{x}_{\mathrm{e}} - \mathbf{x}_{\mathrm{n}}|}, \tag{8.2}$$

where a subscript e or n on ∇ implies the use of derivatives with respect to the components of $\mathbf{x}_{\mathrm{e}}$ or $\mathbf{x}_{\mathrm{n}}$. Remarkably, we can solve this frightening equation exactly. The key step is to introduce six new variables, the components of

$$\mathbf{X} \equiv \frac{m_{\mathrm{e}}\mathbf{x}_e + m_{\mathrm{n}}\mathbf{x}_{\mathrm{n}}}{m_{\mathrm{e}} + m_{\mathrm{n}}} \quad \text{and} \quad \mathbf{r} \equiv \mathbf{x}_{\mathrm{e}} - \mathbf{x}_{\mathrm{n}}. \tag{8.3}$$

$\mathbf{X}$ is the location of the ion's centre of mass, and $\mathbf{r}$ is the vector from the nucleus to the electron. The chain rule yields

$$\frac{\partial}{\partial \mathbf{x}_{\mathrm{e}}} = \frac{\partial \mathbf{X}}{\partial \mathbf{x}_{\mathrm{e}}} \cdot \frac{\partial}{\partial \mathbf{X}} + \frac{\partial \mathbf{r}}{\partial \mathbf{x}_{\mathrm{e}}} \cdot \frac{\partial}{\partial \mathbf{r}} = \frac{m_{\mathrm{e}}}{\mathrm{m_e} + m_{\mathrm{n}}}\frac{\partial}{\partial \mathbf{X}} + \frac{\partial}{\partial \mathbf{r}}. \tag{8.4}$$

When we take the dot product of each side with itself, we find

$$\nabla_{\mathrm{e}}^2 = \left(\frac{m_{\mathrm{e}}}{\mathrm{m_e} + m_{\mathrm{n}}}\right)^2 \nabla_{\mathbf{X}}^2 + \nabla_{\mathbf{r}}^2 + \frac{2m_{\mathrm{e}}}{\mathrm{m_e} + m_{\mathrm{n}}}\frac{\partial^2}{\partial \mathbf{X} \cdot \partial \mathbf{r}}, \tag{8.5a}$$

where the subscripts X and r imply that the operator is to be made up of derivatives with respect to the components of $\mathbf{X}$ or $\mathbf{r}$. Similarly

$$\nabla_{\mathrm{n}}^2 = \left(\frac{m_{\mathrm{n}}}{\mathrm{m}_e + m_{\mathrm{n}}}\right)^2 \nabla_{\mathbf{X}}^2 + \nabla_{\mathbf{r}}^2 - \frac{2m_{\mathrm{n}}}{\mathrm{m_e} + m_{\mathrm{n}}}\frac{\partial^2}{\partial \mathbf{X} \cdot \partial \mathbf{r}}. \tag{8.5b}$$

We now add m_{e}^{-1} times equation (8.5a) to m_{n}^{-1} times equation (8.5b). The mixed derivatives cancel leaving

$$m_{\mathrm{e}}^{-1}\nabla_{\mathrm{e}}^2 + m_{\mathrm{n}}^{-1}\nabla_{\mathrm{n}}^2 = \frac{1}{m_{\mathrm{e}} + m_{\mathrm{n}}}\nabla_{\mathbf{X}}^2 + \frac{1}{\mu}\nabla_{\mathbf{r}}^2, \tag{8.6a}$$

where

$$\mu \equiv \frac{m_{\mathrm{e}} m_{\mathrm{n}}}{m_{\mathrm{e}} + m_{\mathrm{n}}} \tag{8.6b}$$

is called the **reduced mass** of the electron. In the case of hydrogen, when $m_{\mathrm{n}} = m_{\mathrm{p}} = 1685m_{\mathrm{e}}$, the reduced mass differs very little from m_{e} ($\mu = 0.99945m_e$), and in heavier hydrogen-like ions the value of μ lies even closer to m_{e}.

When we use equation (8.6a) to replace $\mathbf{x}_\mathrm{e}$ and $\mathbf{x}_\mathrm{n}$ in equation (8.2) by $\mathbf{r}$ and $\mathbf{X}$, we obtain

$$E\psi = -\frac{\hbar^2}{2(m_\mathrm{e}+m_\mathrm{n})}\nabla^2_{\mathbf{X}}\psi - \frac{\hbar^2}{2\mu}\nabla^2_{\mathbf{r}}\psi - \frac{Ze^2}{4\pi\epsilon_0 r}\psi. \tag{8.7}$$

The right side breaks into two parts: the first term is the Hamiltonian H_K of a free particle of mass $m_\mathrm{e}+m_\mathrm{n}$, while the second and third terms make up the Hamiltonian $H_{\mathbf{r}}$ of a particle of mass μ that is attracted to the origin by an inverse-square law force. Since H_K and $H_{\mathbf{r}}$ commute with one another, there is a complete set of mutual eigenkets. At the end of §6.1 we showed that in these circumstances we can assume that ψ is a product

$$\psi(\mathbf{x}_\mathrm{e},\mathbf{x}_\mathrm{n}) = K(\mathbf{X})\psi_{\mathbf{r}}(\mathbf{r}), \tag{8.8}$$

where

$$-\frac{\hbar^2}{2(m_\mathrm{e}+m_\mathrm{n})}\nabla^2_{\mathbf{X}}K = E_K K \tag{8.9}$$

and

$$-\frac{\hbar^2}{2\mu}\nabla^2_{\mathbf{r}}\psi_{\mathbf{r}} - \frac{Ze^2\psi_{\mathbf{r}}}{4\pi\epsilon_0 r} = E_{\mathbf{r}}\psi_{\mathbf{r}}. \tag{8.10}$$

Here E_K and $E_{\mathbf{r}}$ are two distinct eigenvalues and their sum is the ion's total energy, $E = E_K + E_{\mathbf{r}}$.

From §2.3.3 we know all about the dynamics of a free particle, so equation (8.9) need not detain us. We have to solve equation (8.10). In the interests of simplicity we henceforth omit the subscripts from $\psi_{\mathbf{r}}$ and $E_{\mathbf{r}}$.

Equation (7.69) enables us to write the kinetic energy term in equation (8.10) in terms of the radial momentum operator p_r and the total orbital angular momentum operator L^2. Equation (8.10) is then the eigenvalue equation of the Hamiltonian

$$H \equiv \frac{p_r^2}{2\mu} + \frac{\hbar^2 L^2}{2\mu r^2} - \frac{Ze^2}{4\pi\epsilon_0 r}. \tag{8.11}$$

L^2 commutes with H since the only occurrence in H of the angles θ and ϕ is in L^2 itself. So there is a complete set of mutual eigenstates $|E,l\rangle$ of H and L^2 such that $L^2|E,l\rangle = l(l+1)|E,l\rangle$. For these kets the operator H of equation (8.11) is equivalent to the radial Hamiltonian

$$H_l \equiv \frac{p_r^2}{2\mu} + \frac{l(l+1)\hbar^2}{2\mu r^2} - \frac{Ze^2}{4\pi\epsilon_0 r}. \tag{8.12}$$

This operator is strikingly similar to the radial Hamiltonian defined by equation (7.81) for which we solved the eigenvalue problem in the course of our study of the three-dimensional harmonic oscillator. We use essentially the same technique now, defining the dimensionless ladder operator

$$A_l \equiv \frac{a_0}{\sqrt{2}}\left(\frac{\mathrm{i}}{\hbar}p_r - \frac{l+1}{r} + \frac{Z}{(l+1)a_0}\right), \tag{8.13a}$$

where we have identified the **Bohr radius**[1]

$$a_0 \equiv \frac{4\pi\epsilon_0\hbar^2}{\mu e^2}. \tag{8.13b}$$

---

[1] The physical significance of a_0 is clarified by rewriting equation (8.13b) in the form $e^2/(4\pi\epsilon_0 a_0) = (\hbar/a_0)^2/\mu$. The left side is the electrostatic potential energy at a_0 and the right side is twice the kinetic energy of zero-point motion (§3.1) of a particle whose position has uncertainty $\sim a_0$. For hydrogen $a_0 = 5.29177\times10^{-11}$ m.

The product of A_l with $A_l^\dagger$ is

$$\begin{aligned} A_l^\dagger A_l &= \frac{a_0^2}{2}\left(-\frac{\mathrm{i}}{\hbar}p_r + \frac{Z}{(l+1)a_0} - \frac{l+1}{r}\right)\left(\frac{\mathrm{i}}{\hbar}p_r + \frac{Z}{(l+1)a_0} - \frac{l+1}{r}\right) \\ &= \frac{a_0^2}{2}\left\{\frac{p_r^2}{\hbar^2} + \left(\frac{Z}{(l+1)a_0} - \frac{l+1}{r}\right)^2 + \frac{\mathrm{i}}{\hbar}\left[p_r, \frac{l+1}{r}\right]\right\} \\ &= \frac{a_0^2\mu}{\hbar^2}H_l + \frac{Z^2}{2(l+1)^2}. \end{aligned} \tag{8.14}$$

It is useful to rewrite this result in the form

$$H_l = \frac{\hbar^2}{\mu a_0^2}\left(A_l^\dagger A_l - \frac{Z^2}{2(l+1)^2}\right). \tag{8.15}$$

The commutator of A_l with its Hermitian adjoint $A_l^\dagger$ is

$$\begin{aligned} [A_l, A_l^\dagger] &= \frac{a_0^2}{2}\left[\frac{\mathrm{i}}{\hbar}p_r - \frac{l+1}{r}, -\frac{\mathrm{i}}{\hbar}p_r - \frac{l+1}{r}\right] = \mathrm{i}\frac{a_0^2}{\hbar}\left(\frac{l+1}{r^2}\right)[p_r, r] \\ &= a_0^2\frac{l+1}{r^2} = \frac{a_0^2\mu}{\hbar^2}(H_{l+1} - H_l), \end{aligned} \tag{8.16}$$

where we have used the canonical commutation relation (7.67) and obtained a result very similar to equation (7.86). Using this result and equation (8.15), we can now easily calculate the commutator of H_l with A_l:

$$[A_l, H_l] = \frac{\hbar^2}{\mu a_0^2}[A_l, A_l^\dagger A_l] = \frac{\hbar^2}{\mu a_0^2}[A_l, A_l^\dagger]A_l = (H_{l+1} - H_l)A_l. \tag{8.17}$$

We show that A_l is a ladder operator by multiplying it into both sides of the eigenvalue equation $H_l|E,l\rangle = E|E,l\rangle$:

$$EA_l|E,l\rangle = (H_lA_l + [A_l, H_l])\,|E,l\rangle = H_{l+1}A_l|E,l\rangle. \tag{8.18}$$

This equation states that $A_l|E,l\rangle$ is an eigenket of H_{l+1} with eigenvalue E. That is, A_l transfers energy from the electron's radial motion to its tangential motion. If we repeat this process by multiplying $A_l|E,l\rangle$ by A_{l+1}, and so on, we will eventually arrive at a circular orbit. Let $\mathcal{L}(E)$ denote the l value of this orbit. Then $A_\mathcal{L}$ must annihilate $|E,\mathcal{L}\rangle$ because, if it did not, we would have a state with even greater angular momentum. Thus with equation (8.14) we can write

$$0 = |A_\mathcal{L}|E,\mathcal{L}\rangle|^2 = \langle E,\mathcal{L}|A_\mathcal{L}^\dagger A_\mathcal{L}|E,\mathcal{L}\rangle = \frac{a_0^2\mu}{\hbar^2}E + \frac{Z^2}{2(\mathcal{L}+1)^2}. \tag{8.19}$$

That is,

$$E = -\frac{Z^2\hbar^2}{2\mu a_0^2n^2} = -\frac{Z^2e^2}{8\pi\epsilon_0a_0n^2} = -\frac{Z^2\mu e^4}{2n^2(4\pi\epsilon_0\hbar)^2}, \tag{8.20}$$

where we have defined the **principal quantum number** $n \equiv \mathcal{L}+1$ and the second equality uses the definition (8.13b) of the Bohr radius. The **Rydberg constant** $\mathcal{R}$ is

$$\mathcal{R} \equiv \frac{\hbar^2}{2\mu a_0^2} = \frac{e^2}{8\pi\epsilon_0a_0} = \tfrac{1}{2}\mu\left(\frac{e^2}{4\pi\epsilon_0\hbar}\right)^2 = 13.6056923\,\mathrm{eV}, \tag{8.21}$$

where $\mu = m_\mathrm{e}m_\mathrm{p}/(m_\mathrm{e}+m_\mathrm{p})$ is the reduced mass in the case of hydrogen. The Rydberg constant enables us to give a compact expression for the permitted values of E and l in hydrogen

$$E = -\frac{\mathcal{R}}{n^2} \quad (n = 1, 2, \ldots) \quad ; \quad 0 \le l \le n-1\,. \tag{8.22}$$

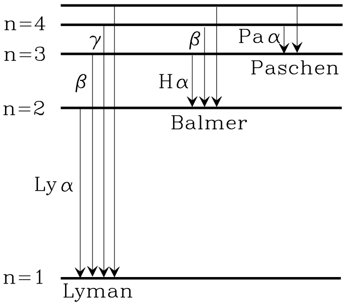


Figure 8.1 Schematic diagram of the Lyman, Balmer and Paschen series of spectral lines in the spectrum of hydrogen.

Henceforth we use n rather than E to label kets and wavefunctions. Thus $|n, l, m\rangle$ is the stationary state of a hydrogen-like ion for the energy given by (8.20) and the stated angular-momentum quantum numbers. the ground state is $|1, 0, 0\rangle$. The energy level immediately above the ground state is four-fold degenerate, being spanned by the states $|2, 0, 0\rangle$, $|2, 1, 0\rangle$ and $|2, 1, \pm 1\rangle$. The second excited energy level is 9-fold degenerate, and so on.

This property of our model hydrogen atom, that it has states with different l but the same energy, is unusual. Atoms with more than one electron have energy levels that depend explicitly on l even when spin and relativity are neglected. When our model of hydrogen is upgraded to include spin and relativity, E becomes weakly dependent on l.

8.1.1 Emission-line spectra

A hydrogen atom may change its value of n to a smaller value n', releasing the liberated energy as a photon of frequency $\nu = (E_n - E_{n'})/h$. Hence the emission spectrum of hydrogen contains lines at the frequencies

$$\nu = \frac{\mathcal{R}}{h}\left(\frac{1}{n'^2} - \frac{1}{n^2}\right). \tag{8.23}$$

The lines associated with a given lower level n' form a series of lines of increasing frequency and decreasing wavelength. The series associated with $n' = 1$ is called the **Lyman series**, the longest-wavelength member of which is the Lyman α line at 121.5 nm, followed by the Lyβ line at 102.5 nm, and so on up to the series limit at 91.2 nm. The series associated with $n' = 2$ is called the **Balmer series** and starts with a line called Hα at 656.2 nm and continues with Hβ at 486.1 nm towards the series limit at 364.6 nm. The series associated with $n' = 3$ is the **Paschen series**, and that associated with $n' = 4$ is the **Brackett** series. Figure 8.1 shows the first three series schematically. Historically the discovery in 1885 by Johann Balmer (1825–1898), a Swiss schoolmaster, that the principal lines in the optical spectrum of hydrogen could be fitted by equation (8.23), was crucial for the development of Niels Bohr's model atom of 1913, which was the precursor of the current quantum theory (Problem 8.2).

Equation (8.20) states that, for given n, the energy of an electron scales as Z^2. For a many–electron atom electromagnetic interactions between the electrons invalidate this scaling. However, it holds to a fair approximation for electrons that have the smallest values of n because these electrons are trapped in the immediate vicinity of the nucleus and their dynamics is largely unaffected by the presence of electrons at larger radii. Henry Moseley (1887–1915) studied the frequencies of X-rays given off when atoms were bombarded by free electrons. He showed[2] that the frequencies of similar spectral lines from different elements seemed to scale with the square of the atomic number. At that time the periodic table was something constructed by chemists that

---

[2] Moseley, H.G.J., 1913, *Phil. Mag.*, **27**, 703. The lines studied by Moseley were associated with transitions $n = 2 \to 1$.

lacked a solid physical foundation. In particular, the atomic numbers of some elements were incorrectly assigned. Moseley's experiments led to the order of cobalt and nickel being reversed, and correctly predicted that elements with atomic numbers 43, 61, 72, 75, 87 and 91 would be discovered.

8.1.2 Radial eigenfunctions

The wavefunctions of hydrogen-like ions are not only important for experiments with atoms and ions that have only one electron, but are also the building blocks from which models of many-electron atoms are built.

The first step in finding any radial eigenfunction for a hydrogen-like ion is to write the equation $A_{n-1}|n, n-1, m\rangle = 0$ (eq. 8.19) as a differential equation for the wavefunction of the circular orbit with angular-momentum quantum number $l = n - 1$. From equations (8.13a) and (7.66) we need to solve

$$\frac{\partial}{\partial r} u_n^{n-1} + \left(-\frac{n-1}{r} + \frac{Z}{na_0}\right) u_n^{n-1} = 0, \tag{8.24}$$

where we have introduced the convention that the subscript on the wavefunction denotes n, while the superscript denotes l. Equation (8.24) is a first-order linear differential equation. Its integrating factor is

$$\exp\left\{\int \mathrm{d}r \left(-\frac{n-1}{r} + \frac{Z}{na_0}\right)\right\} = r^{-(n-1)} \mathrm{e}^{Zr/na_0}, \tag{8.25}$$

so the required eigenfunction is

$$u_n^{n-1}(r) = Cr^{n-1} \mathrm{e}^{-Zr/na_0}, \tag{8.26}$$

where C is a normalising constant. This wavefunction is very similar to our expression (7.95) for the wavefunction of a circular orbit in the three-dimensional harmonic oscillator potential – the only difference is that the Gaussian function has been replaced by a simple exponential. The scale-length in the exponential is $(n/Z)a_0$, so it increases with energy and decreases with the nuclear charge. This makes perfect sense physically because it states that more energetic electrons can go further from a given nucleus, and that nuclei with higher electric charge will bind their (innermost) electrons more tightly.

We determine the normalising constant C in equation (8.26) by multiplying the radial wavefunction by Y_l^m to form the complete eigenfunction, taking the mod square of the result, and integrating it over all space. Bearing in mind that $\mathrm{d}^3\mathbf{x} = r^2 \mathrm{d}r\, \mathrm{d}^2\Omega$ and that $\int \mathrm{d}^2\Omega\, |\mathrm{Y}_l^m|^2 = 1$, we find that C must satisfy

$$\begin{aligned} 1 &= C^2 \int_0^\infty \mathrm{d}r\, r^{2n} \mathrm{e}^{-2Zr/na_0} = C^2 \left(\frac{na_0}{2Z}\right)^{2n+1} \int_0^\infty \mathrm{d}\rho\, \rho^{2n} \mathrm{e}^{-\rho} \\ &= C^2 \left(\frac{na_0}{2Z}\right)^{2n+1} (2n)!, \end{aligned} \tag{8.27}$$

where we have evaluated the integral with the aid of Box 8.1. The correctly normalised radial wavefunction is therefore

$$u_n^{n-1}(r) = \frac{1}{\sqrt{(2n)!}} \left(\frac{2Z}{na_0}\right)^{3/2} \left(\frac{2Zr}{na_0}\right)^{n-1} \mathrm{e}^{-Zr/na_0}. \tag{8.28}$$

We obtain the ground-state radial wavefunction by setting $n = 1$ in equation (8.28):

$$u_1^0(r) = 2 \left(\frac{Z}{a_0}\right)^{3/2} \mathrm{e}^{-Zr/a_0}. \tag{8.29}$$

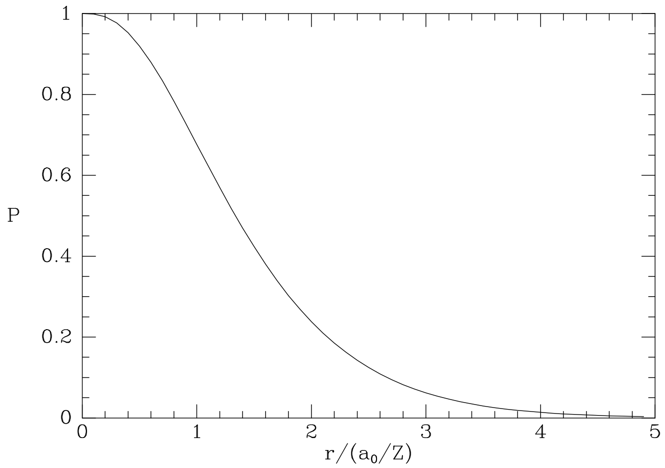


Figure 8.2 The probability of finding the electron of a hydrogen atom that is in its ground state at a radius greater than r. Radii greater than $2a_0/Z$ are classically forbidden.

Box 8.1: The factorial function

We often encounter the integral $\Gamma(\alpha+1) \equiv \int_0^\infty \mathrm{d}t\, t^\alpha \mathrm{e}^{-t}$. Integrating by parts we find that

$$\begin{aligned}\Gamma(\alpha+1) &= -\left[t^\alpha \mathrm{e}^{-t}\right]_0^\infty + \alpha\int_0^\infty \mathrm{d}t\, t^{\alpha-1}\mathrm{e}^{-t}\\ &= \alpha\Gamma(\alpha).\end{aligned}$$

It is easy to check that $\Gamma(1) = 1$. Putting $\alpha = 1$ in the last equation it follows that $\Gamma(2) = 1$. Setting $\alpha = 2$ we find $\Gamma(3) = 2$, and repeating this process we see that for any integer n, $\Gamma(n+1) = n!$. We can use this result to define the factorial function by

$$z! \equiv \Gamma(z+1) = \int_0^\infty \mathrm{d}t\, t^z \mathrm{e}^{-t}. \tag{1}$$

This definition yields a well defined value for $z!$ for any complex number that is not a negative integer, and it coincides with the usual definition of a factorial if z happens to be a non-negative integer.

The complete wavefunction is obtained by multiplying this by $\mathrm{Y}_0^0 = (4\pi)^{-1/2}$. Figure 8.2 shows the probability of finding the electron at a radius greater than r. This reaches $13/\mathrm{e}^4 \simeq 0.24$ at $r = 2a_0/Z$, where the potential energy is equal to the total energy. In classical physics the probability of finding the electron at these radii is zero.

It is interesting to calculate the expectation value of r for circular orbits. We have

$$\begin{aligned}\langle n, n-1, m|r|n, n-1, m\rangle &= \frac{1}{(2n)!}\left(\frac{2Z}{na_0}\right)^3\int_0^\infty \mathrm{d}r\, r^3\left(\frac{2Zr}{na_0}\right)^{2(n-1)}\mathrm{e}^{-2Zr/na_0}\\ &= \frac{na_0}{2Z}\frac{1}{(2n)!}\int_0^\infty \mathrm{d}\rho\, \rho^{2n+1}\mathrm{e}^{-\rho} = n(n+\tfrac{1}{2})\frac{a_0}{Z}.\end{aligned} \tag{8.30}$$

In the classical limit of large n, $\langle r\rangle \simeq n^2 a_0/Z$, so $E \propto 1/n^2 \propto 1/\langle r\rangle$ as classical physics predicts. (One can easily show that classical physics yields the correct proportionality constant.) A very similar calculation shows that

$$\left\langle r^2\right\rangle = n^2(n+1)(n+\tfrac{1}{2})a_0^2/Z^2 = \frac{n+1}{n+\frac{1}{2}}\left\langle r\right\rangle^2, \tag{8.31}$$

so the rms uncertainty in r is $\sqrt{\langle r^2\rangle - \langle r\rangle^2} = \langle r\rangle/\sqrt{2n+1}$. Consequently, as n increases, the uncertainty in r increases as $n^{3/2}$, but the fractional uncertainty in r decreases as $n^{-1/2}$.

Our conclusion that the radius of an atom scales as n^2 implies that an atom with $n \sim 100$ occupies 10^{12} times as much volume as an atom in the ground state. Consequently, only at high-vacuum densities can such highly excited atoms be considered isolated systems. Radio telescopes detect line radiation emitted by hydrogen atoms in the interstellar medium that are reducing their value of n by δn from $n \simeq 100$. The frequency of such a transition is

$$\nu_n = \frac{E_{n+\delta n} - E_n}{h} = \frac{\mathcal{R}}{h}\left(\frac{1}{n^2} - \frac{1}{(n+\delta n)^2}\right) \simeq 6.58\left(\frac{100}{n}\right)^3 \delta n\,\mathrm{GHz}. \tag{8.32}$$

Our analysis of the three-dimensional harmonic oscillator suggests that applications of $A_l^\dagger$ to u_n^{n-1} should generate the wavefunctions u_n^l for $l < n-1$. Here is the proof that they do. Equation (8.16) implies that

$$[A_{l-1}, A_{l-1}^\dagger] = \frac{a_0^2\mu}{\hbar^2}(H_l - H_{l-1}). \tag{8.33}$$

Commuting both sides with $A_{l-1}^\dagger$ yields

$$[H_l, A_{l-1}^\dagger] = [H_{l-1}, A_{l-1}^\dagger] + \frac{\hbar^2}{a_0^2\mu}\Big[[A_{l-1}, A_{l-1}^\dagger], A_{l-1}^\dagger\Big], \tag{8.34}$$

When we use equation (8.14) to eliminate H_{l-1} and expand the inner commutator on the right of equation (8.34) as a difference of products, we obtain

$$[H_l, A_{l-1}^\dagger] = (H_l - H_{l-1})A_{l-1}^\dagger. \tag{8.35}$$

Now when we multiply both sides of $H_l|n,l,m\rangle = E_n|n,l,m\rangle$ by $A_{l-1}^\dagger$ we find

$$E_n A_{l-1}^\dagger|n,l,m\rangle = (H_l A_{l-1}^\dagger + [A_{l-1}^\dagger, H_l])|n,l,m\rangle = H_{l-1}A_{l-1}^\dagger|n,l,m\rangle, \tag{8.36}$$

so $A_{l-1}^\dagger|n,l,m\rangle$ is indeed the hoped-for eigenket of H_{l-1}. Using a result proved in Problem 8.10, we have, in fact, that

$$|n,l-1,m\rangle = \frac{\sqrt{2}}{Z}\left(\frac{1}{l} - \frac{1}{n^2}\right)^{-1/2} A_{l-1}^\dagger|n,l,m\rangle. \tag{8.37}$$

From equations (8.13a) and (7.66) we can write

$$A_{l-1}^\dagger = -\frac{a_0}{\sqrt{2}}\left(\frac{\partial}{\partial r} + \frac{l+1}{r} - \frac{Z}{la_0}\right). \tag{8.38}$$

Setting $l = n-1$ we can apply this operator to u_n^{n-1} to obtain

$$u_n^{n-2}(r) = \text{constant} \times \left(1 - \frac{Zr}{n(n-1)a_0}\right) r^{n-2}\mathrm{e}^{-Zr/na_0}. \tag{8.39}$$

This wavefunction has a node at $r = n(n-1)a_0/Z$. When we apply $A_{n-3}^\dagger$ to this wavefunction to generate u_n^{n-3}, the lowest power of r in the factor that multiplies the exponential will be r^{n-3}, so the exponential will be multiplied by a quadratic in r and the wavefunction will have two nodes. In our study of the three-dimensional harmonic oscillator we encountered the same pattern: the number of nodes in the radial wavefunction increased by one every time $A^\dagger$ decrements the angular momentum and increases the energy of radial motion. The radial eigenfunctions for states with $n \leq 3$ are listed in Table 8.1 and plotted in Figure 8.3.

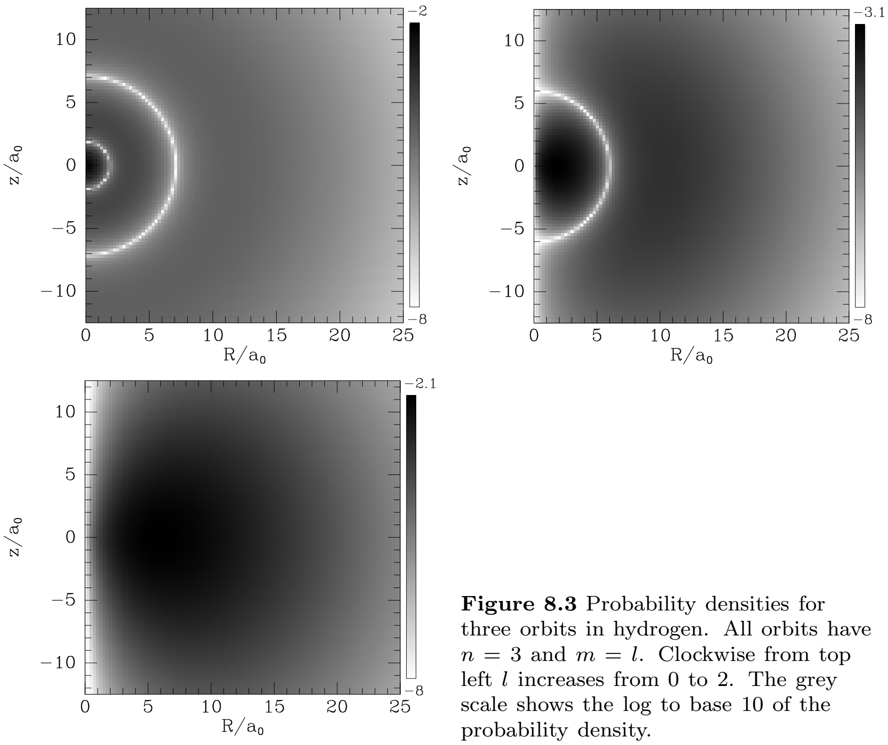


Figure 8.3 Probability densities for three orbits in hydrogen. All orbits have $n = 3$ and $m = l$. Clockwise from top left l increases from 0 to 2. The grey scale shows the log to base 10 of the probability density.

Table 8.1 The first six radial eigenfunctions $u_n^l(r)$ for hydrogen with $a_Z \equiv a_0/Z$. The full wavefunction is $u_n^l(r)\mathrm{Y}_l^m(\theta,\phi)$.

		n	
l	1	2	3
0	$\frac{2e^{-r/a_Z}}{a_Z^{3/2}}$	$\frac{2e^{-r/2a_Z}}{(2a_Z)^{3/2}}\left(1-\frac{r}{2a_Z}\right)$	$\frac{2e^{-r/3a_Z}}{(3a_Z)^{3/2}}\left(1-\frac{2r}{3a_Z}+\frac{2r^2}{27a_Z^2}\right)$
1		$\frac{e^{-r/2a_Z}}{\sqrt{3}(2a_Z)^{3/2}}\frac{r}{a_Z}$	$\frac{2^{5/2}e^{-r/3a_Z}}{9(3a_Z)^{3/2}}\frac{r}{a_Z}\left(1-\frac{r}{6a_Z}\right)$
2			$\frac{2^{3/2}e^{-r/3a_Z}}{\sqrt{27}\sqrt{5}(3a_Z)^{3/2}}\left(\frac{r}{a_Z}\right)^2$

8.1.3 Shielding

The electrostatic potential in which a bound electron moves is never exactly proportional to $1/r$ as we have hitherto assumed. In hydrogen or a single-electron ion the deviations from $1/r$ proportionality are small but measurable. In many electron systems the deviations are large. In all cases the deviations arise because the charge distribution that binds the electron is not confined to a point as we have assumed. First, protons and neutrons have non-zero radii – after all there has to be room for three quarks to move about in there at mildly relativistic speed! Second, even if the nuclear charge *were* confined to a point, the field it generates would not be an inverse-square field because virtual electron-positron pairs make the vacuum a polarisable dielectric medium: in the strong field near the nucleus, the positrons tend to spend their brief lives further from the nucleus, which repels them, than do the electrons. In consequence the charge inside a sphere drawn around the nucleus is slightly smaller than the charge on the nucleus, the charge deficit being small for both very small and very large spheres.

When an atom has more than one electron, these effects are greatly increased since the charge on any electron other than the one whose dynamics we are studying is distributed by quantum uncertainty through the space around the nucleus, so the charge inside a sphere around the nucleus is comparable to the charge on the nucleus when the sphere is very small, but falls to e when the sphere is large.

Phenomena of this type, in which there is a tendency for a charged body to gather oppositely charged bodies around it, are often referred to as 'shielding'. A complete treatment of the action of shielding in even single-electron systems involves quantum field theory and is extremely complex. In this section we modify the results we have obtained so far to explore an idealised model of shielding, which makes it clear how shielding modifies the energy spectrum, and thus the dynamics of atomic species.

The key idea is to replace the atomic number Z in the Hamiltonian with a decreasing function of radius. We adopt

$$Z(r) = Z_0\left(1 + \frac{a}{r}\right), \tag{8.40}$$

where Z_0 and a are adjustable parameters. For $r \gg a$, the nuclear charge tends to a maximally shielded value $Z_0 e$. For $r \simeq a$, the charge is larger by $\sim Z_0 e$. At very small r, the charge diverges, but we anticipate that this unphysical divergence will not have important consequences because the electron is very unlikely to be found at $r \ll a_0/Z$. With this choice for $Z(r)$, the radial Hamiltonian (8.12) becomes

$$H'_l = \frac{p_r^2}{2\mu} + \frac{\{l(l+1) - \beta\}\hbar^2}{2\mu r^2} - \frac{Z_0 e^2}{4\pi\epsilon_0 r}, \tag{8.41a}$$

where

$$\beta \equiv \frac{Z_0 a \mu e^2}{2\pi\epsilon_0\hbar^2}. \tag{8.41b}$$

Because we chose to take the radial dependence of Z to be proportional to $1/r$, we have in the end simply reduced the repulsive centrifugal potential term in the radial Hamiltonian. Let $l'(l)$ be the positive root of the quadratic equation

$$l'(l'+1) = l(l+1) - \beta. \tag{8.42}$$

In general l' will not be an integer. With this definition, H'_l is identical with the Hamiltonian (8.12) of the unshielded case with l' substituted for l and Z_0 replacing Z, that is

$$H'_l = H_{l'(l)}. \tag{8.43}$$

Consequently, the operator $A_{l'}$ that is defined by equation (8.13a) with the same substitutions satisfies (cf. eq. 8.14)

$$A^\dagger_{l'} A_{l'} = \frac{a_0^2\mu}{\hbar^2} H_{l'} + \frac{Z_0^2}{2(l'+1)}. \tag{8.44}$$

Moreover by analogy with equation (8.16) we have

$$[A_{l'}, A^\dagger_{l'}] = \frac{a_0^2\mu}{\hbar^2}(H_{l'+1} - H_{l'}). \tag{8.45}$$

It follows that $A_{l'}$ is a ladder operator

$$E A_{l'}|E, l'\rangle = (H_{l'} A_{l'} + [A_{l'}, H_{l'}])|E, l'\rangle = H_{l'+1} A_{l'}|E, l'\rangle, \tag{8.46}$$

so $A_{l'}|E, l'\rangle = \alpha|E, l'+1\rangle$ is an (unnormalised) eigenket of $H_{l'+1}$ just as in the unshielded case. Applying $A_{l'+1}$ to $|E, l'+1\rangle$ we argue that eventually some maximum value $\mathcal{L}'$ of l' will be reached, at which point $A_{\mathcal{L}'}|E, \mathcal{L}'\rangle = 0$. From the mod square of this equation we conclude that

$$E = -\frac{Z_0^2 e^2}{8\pi\epsilon_0 a_0(\mathcal{L}'+1)^2}, \quad \text{where} \quad \mathcal{L}' = l'(l) + k, \tag{8.47}$$

where k is the number of times we have to apply A to achieve annihilation. Since for $a \neq 0$, $l'(l)$ is not an integer, E is given by the formula (8.20) for the unshielded case with n replaced by a number that is not an integer. Moreover, the energy now depends on l as well as on n, where n is defined to be $1 + l$ plus the number of nodes in the radial wavefunction at $r < \infty$. To see this, consider the effect of increasing our initial value of l by one, and correspondingly decreasing by one the number of times we have to apply $A_{l'}$ to achieve annihilation. In an unshielded atom $l' = l$, so E is unchanged when l is increased and k decreased by unity; we have moved between states with the same value of n. In the shielded case, increasing l by unity does not increment $l'(l)$ by unity, so in equation (8.47) the changes in l' and k do not conspire to hold constant $\mathcal{L}'$. In fact one can show from equation (8.42) that when l increases by one, l' increases by more than one (Problem 8.12), so among states with a given principal quantum number, those with the largest l values have the smallest binding energies. This makes perfect sense physically because it is the eccentric orbits that take the electron close to the nucleus, where the nuclear charge appears greatest.

In 1947 Lamb & Retherford showed[3] that in hydrogen the state $|2,0,0\rangle$ lies 4.4×10^{-6} eV below the states $|2,1,m\rangle$, contrary to naive predictions from the Dirac equation. This **Lamb shift** is due to shielding of the proton by virtual electron-positron pairs in the surrounding vacuum.

8.1.4 Expectation values for r^{-k}

It will prove expedient to have formulae for the expectation value of r^{-k} with hydrogenic wavefunctions and the first three values of k.

The value of $\langle r^{-1}\rangle$ can be obtained from the virial theorem (2.83) since in hydrogen the potential energy is $\propto r^{-1}$. With $\alpha = -1$, equation (2.83) implies that

$$2\langle E|\frac{p^2}{2m}|E\rangle = -\langle E|V|E\rangle. \tag{8.48}$$

On the other hand the expectation of the Hamiltonian yields

$$\langle E|\frac{p^2}{2m}|E\rangle + \langle E|V|E\rangle = E, \tag{8.49}$$

so we have

$$\langle E|V|E\rangle = -\frac{Ze^2}{4\pi\epsilon_0}\langle E|r^{-1}|E\rangle = 2E = -\frac{Z^2e^2}{4\pi\epsilon_0 a_0 n^2}. \tag{8.50}$$

It follows that $\langle r^{-1}\rangle = Z/(n^2 a_0)$.

To obtain $\langle r^{-2}\rangle$ we anticipate a result that we shall prove in §9.1. This relates to what happens when we add a term βH_1 to a system's Hamiltonian, where β is a number and H_1 is an operator. The n^{th} eigenenergy of the complete Hamiltonian then becomes a function of β, and in §9.1 we show that

$$\left.\frac{\mathrm{d}E}{\mathrm{d}\beta}\right|_{\beta=0} = \langle E|H_1|E\rangle. \tag{8.51}$$

We apply this result to a hydrogen-like system with the additional Hamiltonian

$$H_1 = -\frac{\hbar^2}{2\mu r^2}. \tag{8.52}$$

In the last subsection we showed that the exact eigenvalues of this system are given by equation (8.47). Differentiating the eigenvalues with respect to β and using equation (8.51) we find

$$\frac{\hbar^2}{2\mu}\langle E|r^{-2}|E\rangle = \left.\frac{\mathrm{d}}{\mathrm{d}\beta}\right|_{\beta=0} \frac{Z^2e^2}{8\pi\epsilon_0 a_0(l'+k)^2} = -\frac{Z^2e^2}{4\pi\epsilon_0 a_0(l+k)^3}\frac{\mathrm{d}l'}{\mathrm{d}\beta} \tag{8.53}$$

---

[3] W.E. Lamb & R.C. Retherford, *Phys. Rev.* **72**, 241

From equation (8.42) we have $\mathrm{d}l'/\mathrm{d}\beta = -1/(2l+1)$, so

$$\langle E|r^{-2}|E\rangle = \frac{Z^2e^2\mu}{2\pi\epsilon_0 a_0\hbar^2 n^3(2l+1)} = \frac{Z^2}{a_0^2 n^3(l+\frac{1}{2})}, \tag{8.54}$$

where the last equality uses the definition (8.13b) of the Bohr radius.

We determine $\left\langle r^{-3}\right\rangle$ by considering the expectation value of the commutator $[H,p_r]$. As we saw in §2.2.1, in a stationary state the commutator with H of any observable vanishes. Hence with equation (8.12) we can write

$$0 = \langle E|[H,p_r]|E\rangle = \frac{l(l+1)\hbar^2}{2\mu}\langle E|[r^{-2},p_r]|E\rangle - \frac{Ze^2}{4\pi\epsilon_0}\langle E|[r^{-1},p_r]|E\rangle \tag{8.55}$$

Using the canonical commutation relation $[r,p_r] = \mathrm{i}\hbar$ [equation (7.67)] to evaluate the commutators in this expression, and the value of $\left\langle r^{-2}\right\rangle$ that we have just established, we find

$$\langle E|r^{-3}|E\rangle = \frac{Z^3}{a_0^3 n^3 l(l+1)(l+\frac{1}{2})}. \tag{8.56}$$

The three values of $\left\langle r^{-k}\right\rangle$ that we have calculated conform to a pattern. First the basic atomic scale a_0/Z is raised to the $-k^{\text{th}}$ power. Then there is a product of $2k$ quantum numbers on the bottom, reflecting the tendency for the atom's size to grow as n^2. Finally, as k increases, the number of factors of l increases from zero to three, reflecting the growing sensitivity of $\left\langle r^{-3}\right\rangle$ to orbital eccentricity.

8.2 Fine structure and beyond

The model of hydrogen-like ions that we developed in the last section is satisfying and useful, but it is far from complete. We now consider some of the physics that is neglected by this model.

In §2.3.5 we saw that when a particle that moves in an inverse-square force field is in a stationary state, the expectation value of its kinetic energy, classically $\frac{1}{2}mv^2$, is equal in magnitude but opposite in sign to its total energy. Equation (8.20) is an expression for the ground-state energy of an electron in a hydrogen-like ion. When we equate the absolute value of this expression to $\frac{1}{2}m_\mathrm{e}v^2$, we find that the ratio of v to the speed of light c is

$$\frac{v}{c} = \alpha Z, \tag{8.57}$$

where the dimensionless **fine structure constant** is defined to be

$$\alpha \equiv \frac{e^2}{4\pi\epsilon_0\hbar c} \simeq \frac{1}{137}. \tag{8.58}$$

Since relativistic corrections tend to be $\mathrm{O}(v^2/c^2)$, it follows that in hydrogen relativistic corrections to the results we have derived may be expected to be several parts in 10^5, but these corrections, being proportional to Z^2, will exceed 10% by the middle of the periodic table.

For future reference we note that with the reduced mass μ approximated by m_e, equation (8.13b) for the Bohr radius can be written

$$a_0 = \frac{\hbar}{\alpha m_\mathrm{e} c} = \frac{\lambda_{\text{Compton}}}{2\pi\alpha}, \tag{8.59}$$

where we have identified the electron's **Compton wavelength** $h/m_\mathrm{e}c$ (the wavelength of a photon that has energy $m_\mathrm{e}c^2$). When we use this expression to eliminate a_0 from equation (8.20), we find that the energy levels of a hydrogen-like ion are

$$E = -\frac{Z^2\alpha^2}{2n^2}m_\mathrm{e}c^2, \quad \text{so} \quad \mathcal{R} = \tfrac{1}{2}\alpha^2 m_\mathrm{e}c^2. \tag{8.60}$$

8.2.1 Spin-orbit coupling

Magnetism is a relativistic correction to electrostatics in the sense that a particle that is moving with velocity $\mathbf{v}$ in an electric field $\mathbf{E}$ experiences a magnetic field

$$\mathbf{B} = \frac{1}{c^2}\mathbf{v} \times \mathbf{E}. \tag{8.61}$$

If the particle has a magnetic dipole moment $\boldsymbol{\mu}$, it experiences a torque $\mathbf{G} = \boldsymbol{\mu} \times \mathbf{B}$ that will cause its spin $\mathbf{S}$ to precess. In the particle's rest frame the classical equation of motion of $\mathbf{S}$ is

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{1}{\hbar}\boldsymbol{\mu} \times \mathbf{B}, \tag{8.62}$$

where $\hbar$ appears only because $\mathbf{S}$ is the dimensionless spin obtained by dividing the angular momentum by $\hbar$. We assume that the magnetic moment $\boldsymbol{\mu}$ is proportional to the dimensionless spin vector $\mathbf{S}$ and write the proportionality

$$\boldsymbol{\mu} = \frac{gQ\hbar}{2m_0}\mathbf{S}, \tag{8.63}$$

where g is the dimensionless **gyromagnetic ratio**, and Q and m_0 are the particle's charge and rest mass. In the case of an electron $g = 2.002$, a value which is correctly predicted by relativistic quantum electrodynamics, and the dimensional factor is defined to be the **Bohr magneton**

$$\mu_\mathrm{B} \equiv \frac{e\hbar}{2m_\mathrm{e}} = 9.27 \times 10^{-24}\,\mathrm{J\,T^{-1}}. \tag{8.64}$$

With this notation, our rest-frame equation of motion (8.62) becomes

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{gQ}{2m_0}\mathbf{S} \times \mathbf{B}. \tag{8.65}$$

The non-zero value of the right side of this classical equation of motion for $\mathbf{S}$ implies that there is a spin-dependent term in the particle's Hamiltonian since the operator $\mathbf{S}$ commutes with all spatial operators (§7.4) and the right side of the classical equation of motion (8.65) is proportional to the expectation of $[\mathbf{S}, H]$ (cf. eq. 2.57). We want to determine what this term in H is.

Energy is not a relativistic invariant – it is physically obvious that observers who move relative to one another assign different energies to a given system. Consequently, when they do quantum mechanics they use different Hamiltonians. We need the Hamiltonian that governs the dynamics of the reduced particle in the rest frame of the atom's centre of mass. So we have to transform the equation of motion (8.65) to this frame. This is a tricky business because the reduced particle is accelerating, so the required Lorentz transformation is time-dependent. Given the delicacy of the required transformation, it is advisable to work throughout with explicitly Lorentz 'covariant' quantities, which are explained in Appendix D. In Appendix E these are used to show that in a frame of reference in which the electron is moving, equation (8.65) becomes

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{Q}{2m_0c^2}\left(-\frac{\hbar}{m_0r}\frac{\mathrm{d}\Phi}{\mathrm{d}r}\mathbf{S} \times \mathbf{L} + 2c^2\mathbf{S} \times \mathbf{B}\right). \tag{8.66}$$

It is straightforward to demonstrate (Problem 8.13) from equation (2.34) that this classical equation of motion of the spin $\mathbf{S}$ of an electron (which has charge $Q = -e$) arises if we introduce into the quantum-mechanical Hamiltonian (8.1) two spin-dependent terms, namely the **spin-orbit Hamiltonian**

$$H_\mathrm{SO} \equiv -\frac{\mathrm{d}\Phi}{\mathrm{d}r}\frac{e\hbar^2}{2rm_\mathrm{e}^2c^2}\mathbf{S} \cdot \mathbf{L}, \tag{8.67}$$

and the **Zeeman spin Hamiltonian**

$$H_{\rm ZS} \equiv \frac{e\hbar}{m_{\rm e}}\mathbf{S}\cdot\mathbf{B}. \tag{8.68}$$

The Zeeman spin Hamiltonian is just $\boldsymbol{\mu}\cdot\mathbf{B}$ with equation (8.63) used to replace the magnetic moment operator by the spin operator. Interestingly, the spin-orbit Hamiltonian is a factor two smaller than $\boldsymbol{\mu}\cdot\mathbf{B}$ with $\boldsymbol{\mu}$ replaced in the same way and equation (8.61) used to relate $\mathbf{B}$ to the electric field in which the electron is moving. In the 1920s the experimental data clearly required this factor of two difference in the spin Hamiltonians, but its origin puzzled the pioneers of the subject until, in 1927, L.T. Thomas[4] showed that it is a consequence of the fact that the electron's rest frame is accelerating (Appendix E). If no torque is applied to a gyro, it does not precess in its instantaneous rest frame. But if the direction of the gyro's motion is changing relative to some inertial frame, the sequence of Lorentz transformations that are required to transform the spin vector into the inertial frame causes the spin to precess in the inertial frame. This apparent precession of an accelerated gyro is called **Thomas precession**.

In a single-electron system such as hydrogen, $\Phi = Ze/(4\pi\epsilon_0 r)$, so

$$H_{\rm SO} = \frac{Z\alpha\hbar^3}{2m_{\rm e}^2cr^3}\mathbf{S}\cdot\mathbf{L}, \tag{8.69}$$

where the fine-structure constant (8.58) has been used to absorb the $4\pi\epsilon_0$. Since the coefficient in front of the operator $\mathbf{S}\cdot\mathbf{L}$ is positive, spin-orbit coupling lowers the energy when the spin and orbital angular momenta are antiparallel.

The operator $\mathbf{S}\cdot\mathbf{L}$ in equation (8.69) is most conveniently written

$$\mathbf{S}\cdot\mathbf{L} = \tfrac12((\mathbf{L}+\mathbf{S})^2 - L^2 - S^2) = \tfrac12(J^2 - L^2 - S^2), \tag{8.70}$$

so $H_{\rm SO}$ is diagonal in a basis made up of mutual eigenkets of J^2, L^2 and S^2. In §7.5 we constructed such mutual eigenkets from the eigenkets of S^2, S_z, L^2, and L_z. $\mathbf{S}\cdot\mathbf{L}$ annihilates states with quantum number $l = 0$ because then $j = s$. Hence, there is no spin-orbit coupling in the ground state of hydrogen. In any excited state, $l > 0$ is permitted, and from §7.5.2 we know that the possible values of j are $l \pm \frac12$. The associated eigenvalues of the operator on the right of equation (8.70) are readily found to be

$$\tfrac12\{j(j+1) - l(l+1) - \tfrac34\} = \begin{cases} \tfrac12 l & \text{for } j = l + \tfrac12, \\ -\tfrac12(l+1) & \text{for } j = l - \tfrac12. \end{cases} \tag{8.71}$$

Although $\mathbf{S}\cdot\mathbf{L}$ commutes with the gross-structure Hamiltonian $H_{\rm GS}$ (eq. 8.1), the other operator in $H_{\rm SO}$, namely r^{-3}, does not. So the eigenkets of $H_{\rm GS} + H_{\rm SO}$ will differ (subtly) from the eigenkets we have found. In §9.1 we shall show that in these circumstances the change in the energy of a stationary state can be estimated by replacing the operator by its expectation value. Equation (8.56) gives this value, and, inserting this with our results for the spin operators, yields energy shifts

$$\Delta E \equiv \langle n,l,m|H_{\rm SO}|n,l,m\rangle \simeq K_n \Big/ \begin{cases} (l+1)(l+\tfrac12) & \text{for } j = l+\tfrac12 \\ -l(l+\tfrac12) & \text{for } j = l-\tfrac12 \end{cases} \quad (l>0), \tag{8.72a}$$

where

$$K_n \equiv \frac{Z^4\alpha\hbar^3}{4a_0^3m_{\rm e}^2cn^3} = \frac{Z^4\alpha^4}{4n^3}m_{\rm e}c^2, \tag{8.72b}$$

---

[4] *Phil. Mag.* **3**, 1 (1927)

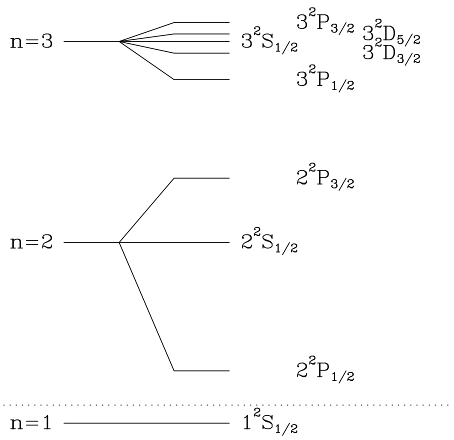


Figure 8.4 The fine structure of a hydrogen-like ion with $Z = 200$ that is predicted by equation (8.72a). The dotted line denotes a break in the energy scale so that the ground state can be included.

and the second equality uses equation (8.59) for a_0. The difference between the energies of states with $j = l \pm \frac{1}{2}$ is

$$E_{l+1/2} - E_{l-1/2} = \frac{2K_n}{l(l+1)}. \tag{8.72c}$$

For $n = 1$ the fine-structure energy scale K_n is smaller than the gross-structure energy (8.60) by a factor $Z^2/2\alpha^2$ that rises from parts in 10^5 for hydrogen to more than 10% by the middle of the periodic table.[5] In hydrogen fine-structure is largest in the $n = 2$, $l = 1$ level, which is split into $j = \frac{3}{2}$ and $j = \frac{1}{2}$ sublevels. According to equations (8.72c), these sublevels are separated by $K_2 = 4.53 \times 10^{-5}\,\text{eV}$, while the measured shift is $4.54 \times 10^{-5}\,\text{eV}$.

Figure 8.4 shows the prediction of equation (8.72a) for the energy levels of a hydrogen-like ion with $Z = 200$. With this unrealistically large value of Z the fine structure for $n = 2$ has comparable magnitude to the gross-structure difference between the $n = 2$ and $n = 3$ levels. The levels in this figure are labelled in an obscure notation that is traditional in atomic physics and more fully explaine in Box 10.2. The value of n appears first, followed by one of the letters S, P, D, F to denote $l = 0, 1, 2, 3$, respectively.[6] The value of j appears as a subscript to the letter, and the value of $2s + 1$ (here always 2) appears before the letter as a superscript. So the level $2^2P_{3/2}$ has $n = 2$, $s = 1/2$, $l = 1$, and $j = \frac{3}{2}$. From Figure 8.4 we see that states in which j is less than l (because the electron's spin and orbital angular momenta are antiparallel) are predicted to have lower energies than the corresponding states in which the two angular momenta are aligned. The spin–orbit interaction vanishes by symmetry for $s = 0$ but otherwise at fixed n the magnitude of the effect decreases with increasing angular momentum because the electron's top speed on a nearly circular orbit is smaller than on an eccentric orbit, so relativistic effects are largest on eccentric orbits.

Equation (8.72a) suggests that states that differ in l but not j should have different energies, whereas they do in fact have extremely similar energies. For example, the $2^2S_{1/2}$ state lies $4.383 \times 10^{-6}\,\text{eV}$ above the $2^2P_{1/2}$ state, while equation (8.72a) implies that this energy difference should be $\frac{3}{2}K_2 = 6.79 \times 10^{-5}\,\text{eV}$. This discrepancy arises because the spin-orbit Hamiltonian does not provide a complete description to order α^4 of relativistic corrections to the electrostatic Hamiltonian. Actually, additional corrections shift the energy of the $2^2S_{1/2}$ states into close alignment with the energy of

---

[5] Naturally, the fine-structure constant owes its name to its appearance in this ratio of the fine-structure and gross energies.

[6] These letters are a shorthand for a description of spectral lines that later were found to involve the various l values: **s**harp, **p**rincipal, **d**iffuse, **f**aint.

the $2^2P_{1/2}$ states.[7] However, in atoms with more than one electron, the electrostatic repulsion between the electrons shifts the energy of the $2^2S_{1/2}$ states downwards by much larger amounts. These electrostatic corrections are hard to calculate accurately, so the much smaller relativistic corrections are not interesting, experimentally, and the quantities of interest are difference in energy between states with the same values of l but different j. These difference are correctly given by equation (8.72a).

Relativistic quantum electrodynamics is in perfect agreement with measurements of hydrogen. It uses the Dirac equation rather than classically-inspired corrections to the electrostatic Hamiltonian. We have devoted significant space to deriving the spin-orbit Hamiltonian not because it plays a role in hydrogen, but because it becomes important as one proceeds down the periodic table. The other relativistic corrections also become large by the middle of the periodic table, but outside hydrogen their effects are so masked by electron-electron interactions that they are of little practical importance and we shall not discuss them in this book.

8.2.2 Hyperfine structure

A proton is a charged spin-half particle, so like an electron it has a magnetic moment. By analogy with the definition of the Bohr magneton (eq. 8.64), we define the **nuclear magneton** to be

$$\mu_{\rm p} \equiv \frac{e\hbar}{2m_{\rm p}} = 5.05\times10^{-27}\,{\rm J\,T^{-1}}. \tag{8.74}$$

In terms of $\mu_{\rm p}$, the magnetic-moment operator of the proton is

$$\boldsymbol{\mu} = g_{\rm p}\mu_{\rm p}\mathbf{S}_{\rm p}, \tag{8.75}$$

where $g_{\rm p} = 5.58$ and $\mathbf{S}_{\rm p}$ is proton's spin operator, so the proton's magnetic moment is smaller than that of an electron by a factor $2.79m_{\rm e}/m_{\rm p} \simeq 1.5\times10^{-3}$.

The electron in a hydrogen atom can create a magnetic field at the location of the proton in two ways: as a moving charge, it generates a current, and it has its intrinsic magnetic moment, so its probability distribution $|\psi(\mathbf{x})|^2$ is a distribution of magnetic dipoles that will generate a magnetic field just as iron does in a bar magnet.

The ground level of hydrogen is a particularly simple case because in this state the electron has no orbital angular momentum, so it generates a magnetic field exclusively through its dipole moment. The magnetic vector potential distance r from a magnetic dipole $\boldsymbol{\mu}_{\rm e}$ is

$$\mathbf{A} = \frac{\mu_0}{4\pi}\frac{\boldsymbol{\mu}_{\rm e}\times\mathbf{r}}{r^3} = \frac{\mu_0}{4\pi}\nabla\times\left(\frac{\boldsymbol{\mu}_{\rm e}}{r}\right). \tag{8.76}$$

The magnetic field is $\mathbf{B} = \nabla\times\mathbf{A}$, so the **hyperfine-structure Hamiltonian** for the ground state is

$$H_{\rm HFS} = \boldsymbol{\mu}_{\rm p}\cdot\mathbf{B} = \frac{\mu_0}{4\pi}\boldsymbol{\mu}_{\rm p}\cdot\nabla\times\left\{\nabla\times\left(\frac{\boldsymbol{\mu}_{\rm e}}{r}\right)\right\}. \tag{8.77}$$

---

[7] In the lowest order of relativistic quantum electrodynamics, the energy of a hydrogen atom depends on only n and j: the Dirac equation predicts

$$E = -\frac{\mathcal{R}}{n^2}\left\{1-\frac{\alpha^2Z^2}{n^2}\left(\frac{3}{4}-\frac{n}{j+\frac{1}{2}}\right)\right\}. \tag{8.73}$$

Thus the $2^2S_{1/2}$ states are predicted to have the same energy as the $2^2P_{1/2}$ states. The measured Lamb shift between these states arises in the next order as a consequence of polarisation of the vacuum, as described in §8.1.3.

Until $H_{\rm HFS}$ is included in the atom's Hamiltonian, the atom's lowest energy level is degenerate because the spins of the electron and the proton can be combined in a number of different ways. To proceed further we need to evaluate the matrix elements obtained by squeezing $H_{\rm HFS}$ between states that form a basis for the ground-level states. The natural basis to use is made up of the states $|j=0\rangle$ and $|j=1,m\rangle$ for $m=-1,0,1$ that can be constructed by adding two spin-half systems (§7.5.1). In Appendix F we show that the resulting matrix elements are

$$\begin{aligned}\langle\psi,s|H_{\rm HFS}|\psi,s'\rangle &= \frac{2\mu_0}{3}|\psi(0)|^2\langle s|\boldsymbol{\mu}_{\rm p}\cdot\boldsymbol{\mu}_{\rm e}|s'\rangle\\ &= -\frac{2\mu_0}{3}|\psi(0)|^2 g_{\rm p}\mu_{\rm p}2\mu_{\rm B}\langle s|\mathbf{S}_{\rm p}\cdot\mathbf{S}_{\rm e}|s'\rangle,\end{aligned} \tag{8.78}$$

where we have replaced the magnetic moment operators by the appropriate multiples of the spin operators. From our discussion of the spin-orbit Hamiltonian (8.69), which is also proportional to the dot product of two angular-momentum operators, we know that the eigenstates of the total angular momentum operators are simultaneously eigenstates of $\mathbf{S}_{\rm p}\cdot\mathbf{S}_{\rm e}$ with eigenvalues $\frac{1}{2}\{j(j+1)-\frac{3}{4}-\frac{3}{4}\}$, so in this basis the off-diagonal matrix elements vanish and the diagonal ones are

$$\langle\psi,j,m|H_{\rm HFS}|\psi,j,m\rangle = -\frac{2\mu_0}{3}|\psi(0)|^2 g_{\rm p}\mu_{\rm p}\mu_{\rm B}\{j(j+1)-\tfrac{3}{2}\}. \tag{8.79}$$

In §9.1 we shall show that the diagonal matrix elements provide a good estimates of the amount by which $H_{\rm HFS}$ shifts the energies of the stationary states of the gross-structure Hamiltonian.

The total angular-momentum quantum number of the atom can be $j=0$ or $j=1$ and the two possible values of the curly bracket above differ by two. Equation (8.29) gives $|\psi(0)|^2=1/(\pi a_0^3)$, so the energies of these levels differ by

$$\Delta E = \frac{4\mu_0}{3\pi a_0^3}5.58\mu_{\rm p}\mu_{\rm B} = 5.88\times10^{-6}\,{\rm eV}. \tag{8.80}$$

The lower level, having $j=0$, is non-degenerate, while the excited state is three-fold degenerate. Transitions between these levels give rise to radiation of frequency 1.420 405 7518 GHz. To obtain perfect agreement between this most accurately measured frequency and equation (8.80), it is necessary to change the gyromagnetic ratio of the electron from the value of 2 that we have adopted to the value 2.002319 . . . that is predicted by quantum electrodynamics. The agreement between theory and experiment is then impressive.

The hyperfine line of hydrogen provides the most powerful way of tracing diffuse gas in interstellar and intergalactic space. Radiation at this frequency can propagate with little absorption right through clouds of dust and gas that absorb optical radiation. Consequently it was in radiation at 1.4 GHz that the large-scale structure of our own galaxy was first revealed in the 1950s. The line is intrinsically very narrow with the consequence that the temperature and radial velocity of the hydrogen that emits the radiation can be accurately measured from the Doppler shift and broadening in the observed spectral line. The existence of 1.4 GHz line radiation from our galaxy was predicted theoretically by H.C. van de Hulst as part of his doctoral work in Nazi-occupied Utrecht. In 1951 groups in the USA and Australia and the Netherlands, detected the line almost simultaneously. The Dutch group used German radar antenna left over from the war.

Problems

8.1 Some things about hydrogen's gross structure that it's important to know (ignore spin throughout):

a) What quantum numbers characterise stationary states of hydrogen?
b) What combinations of values of these numbers are permitted?
c) Give the formula for the energy of a stationary state in terms of the Rydberg $\mathcal{R}$. What is the value of $\mathcal{R}$ in eV?
d) How many stationary states are there in the first excited level and in the second excited level?
e) What is the wavefunction of the ground state?
f) Write down an expression for the mass of the reduced particle.
g) We can apply hydrogenic formulae to any two charged particles that are electrostatically bound. How does the ground-state energy then scale with (i) the mass of the reduced particle, and (ii) the charge Ze on the ion? (iii) How does the radial scale of the system scale with Z?

8.2 In the Bohr atom, electrons move on classical circular orbits that have angular momenta $l\hbar$, where $l = 1, 2, \ldots$. Show that the radius of the first Bohr orbit is a_0 and that the model predicts the correct energy spectrum. In fact the ground state of hydrogen has zero angular momentum. Why did Bohr get correct answers from an incorrect hypothesis?

8.3 Show that the speed of a classical electron in the lowest Bohr orbit (Problem 8.2) is $v = \alpha c$, where $\alpha = e^2/4\pi\epsilon_0\hbar c$ is the fine-structure constant. What is the corresponding speed for a hydrogen-like Fe ion (atomic number $Z = 26$)?

8.4 Show that Bohr's hypothesis (that a particle's angular momentum must be an integer multiple of $\hbar$), when applied to the three-dimensional harmonic oscillator, predicts energy levels $E = l\hbar\omega$ with $l = 1, 2, \ldots$. Is there an experiment that would falsify this prediction?

8.5 Show that the electric field experienced by an electron in the ground state of hydrogen is of order $5 \times 10^{11}\,\mathrm{V\,m^{-1}}$. Can comparable macroscopic fields be generated in the laboratory?

8.6 Positronium consists of an electron and a positron (both spin-half and of equal mass) in orbit around one another. What are its energy levels? By what factor is a positronium atom bigger than a hydrogen atom?

8.7 The emission spectrum of the He^+ ion contains the Pickering series of spectral lines that is analogous to the Lyman, Balmer and Paschen series in the spectrum of hydrogen.

Balmer $i = 1, 2, \ldots$	0.456806	0.616682	0.690685	0.730884
Pickering $i = 1, 3, \ldots$	0.456987	0.616933	0.690967	0.731183

The table gives the frequencies (in 10^{15} Hz) of the first four lines of the Balmer series and the first four odd-numbered lines of the Pickering series. The frequencies of alternate lines in the Pickering series are almost coincident with the frequencies of lines of the Balmer series. Explain this finding. Provide a quantitative explanation of the small offset between these nearly coincident lines in terms of the reduced mass of the electron in the two systems.

8.8 Tritium, ^{3}H, is highly radioactive and decays with a half-life of 12.3 years to ^{3}He by the emission of an electron from its nucleus. The electron departs with 16 keV of kinetic energy. Explain why its departure can be treated as sudden in the sense that the electron of the original tritium atom barely moves while the ejected electron leaves.

Calculate the probability that the newly-formed ^{3}He atom is in an excited state. Hint: evaluate $\langle 100; Z = 2|100; Z = 1\rangle$.

8.9* A spherical potential well is defined by

$$V(r) = \begin{cases} 0 & \text{for } r < a \\ V_0 & \text{otherwise,} \end{cases} \tag{8.81}$$

where $V_0 > 0$. Consider a stationary state with angular-momentum quantum number l. By writing the wavefunction $\psi(\mathbf{x}) = R(r)\mathrm{Y}_l^m(\theta, \phi)$ and using $p^2 = p_r^2 + \hbar^2 L^2/r^2$, show that the state's radial wavefunction $R(r)$ must satisfy

$$-\frac{\hbar^2}{2m}\left(\frac{\mathrm{d}}{\mathrm{d}r} + \frac{1}{r}\right)^2 R + \frac{l(l+1)\hbar^2}{2mr^2} R + V(r)R = ER. \tag{8.82}$$

Show that in terms of $S(r) \equiv rR(r)$, this can be reduced to

$$\frac{\mathrm{d}^2 S}{\mathrm{d}r^2} - l(l+1)\frac{S}{r^2} + \frac{2m}{\hbar^2}(E - V)S = 0. \tag{8.83}$$

Assume that $V_0 > E > 0$. For the case $l = 0$ write down solutions to this equation valid at (a) $r < a$ and (b) $r > a$. Ensure that R does not diverge at the origin. What conditions must S satisfy at $r = a$? Show that these conditions can be simultaneously satisfied if and only if a solution can be found to $k \cot ka = -K$, where $\hbar^2 k^2 = 2mE$ and $\hbar^2 K^2 = 2m(V_0 - E)$. Show graphically that the equation can only be solved when $\sqrt{2mV_0}\, a/\hbar > \pi/2$. Compare this result with that obtained for the corresponding one-dimensional potential well.

The deuteron is a bound state of a proton and a neutron with zero angular momentum. Assume that the strong force that binds them produces a sharp potential step of height V_0 at interparticle distance $a = 2 \times 10^{-15}$ m. Determine in MeV the minimum value of V_0 for the deuteron to exist. Hint: remember to consider the dynamics of the *reduced* particle.

8.10* Given that the ladder operators for hydrogen satisfy

$$A_l^\dagger A_l = \frac{a_0^2 \mu}{\hbar^2} H_l + \frac{Z^2}{2(l+1)^2} \quad \text{and} \quad [A_l, A_l^\dagger] = \frac{a_0^2 \mu}{\hbar^2}(H_{l+1} - H_l)$$

where H_l is the Hamiltonian for angular-momentum quantum number l, show that

$$A_{l-1} A_{l-1}^\dagger = \frac{a_0^2 \mu}{\hbar^2} H_l + \frac{Z^2}{2l^2}. \tag{8.84}$$

Hence show that

$$A_{l-1}^\dagger |E, l\rangle = \frac{Z}{\sqrt{2}}\left(\frac{1}{l^2} - \frac{1}{n^2}\right)^{1/2} |E, l-1\rangle, \tag{8.85}$$

where n is the principal quantum number. Explain the physical meaning of this equation and its use in setting up the theory of the hydrogen atom.

8.11 Show that for hydrogen the matrix element $\langle 200|z|210\rangle = -3a_0$. In §9.1.2 it will be important that this matrix element is non-zero.

8.12* From equation (8.42) show that $l' + \frac{1}{2} = \sqrt{(l + \frac{1}{2})^2 - \beta}$ and that the increment Δ in l' when l is increased by one satisfies $\Delta^2 + \Delta(2l'+1) = 2(l+1)$. By considering the amount by which the solution of this equation changes when l' changes from l as a result of β increasing from zero to a small number, show that

$$\Delta = 1 + \frac{2\beta}{4l^2 - 1} + \mathrm{O}(\beta^2). \tag{8.86}$$

Explain the physical significance of this result.

8.13 Show that for Ehrenfest's theorem yields equation (8.66) with $\mathbf{B} = 0$ as the classical equation of motion of the vector $\mathbf{S}$ that is implied by the spin–orbit Hamiltonian (8.67).

9

Perturbation theory

It is rarely possible to solve exactly for the dynamics of a system of experimental interest. In these circumstances we use some kind of approximation to tweak the solution to some model system that is as close as possible to the system of interest and yet is simple enough to have analytically solvable dynamics. That is, we treat the difference between the experimental system and the model system as a 'perturbation' of the model. Perturbation theory in this sense was an important part of mathematical physics before quantum mechanics appeared on the scene – in fact the development of Hamiltonian mechanics was driven by people who were using perturbation theory to understand the dynamics of the solar system. Interestingly, while perturbation theory in classical mechanics remains an eclectic branch of knowledge that is understood only by a select few, perturbation theory in quantum mechanics is a part of main-stream undergraduate syllabuses. There are two reasons for this. First, analytically soluble models are even rarer in quantum than in classical physics, so more systems have to be modelled approximately. Second, in quantum mechanics perturbation theory is a good deal simpler and works rather better than in classical mechanics.

9.1 Time-independent perturbations

Let H be the Hamiltonian of the experimental system and H_0 the Hamiltonian of the model system for which we have already solved the eigenvalue problem. We hope that $\Delta \equiv H - H_0$ is small and define

$$H_\beta = H_0 + \beta\Delta. \tag{9.1}$$

We can think of H_β as the Hamiltonian of an apparatus that has a knob on it labelled 'β'; when the knob is turned to $\beta = 0$, the apparatus is the model system, and as the knob is turned round to $\beta = 1$, the apparatus is gradually deformed into the system of experimental interest.

We seek the eigenkets $|E\rangle$ and eigenvalues E of H_β as functions of β. Since the Hamiltonian of the apparatus is a continuous function of β, we conjecture that the $|E\rangle$ and E are continuous functions of β too. In fact, we conjecture that they are analytic functions[1] of β so they can be expanded as power series

$$|E\rangle = |a\rangle + \beta|b\rangle + \beta^2|c\rangle + \cdots \quad ; \quad E = E_a + \beta E_b + \beta^2 E_c + \cdots, \tag{9.2}$$

where $|a\rangle$, $|b\rangle$, etc., are states to be determined and E_a, E_b, etc., are appropriate numbers. When we plug our conjectured forms (9.2) into the eigenvalue equation $H|\psi\rangle = E|\psi\rangle$, we have

$$(H_0+\beta\Delta)\big(|a\rangle+\beta|b\rangle+\beta^2|c\rangle+\big) = \big(E_a+\beta E_b+\beta^2 E_c+\big)\ \big(|a\rangle+\beta|b\rangle+\beta^2|c\rangle+\big). \tag{9.3}$$

Since we require the equality to hold for any value of β, we can equate the coefficient of every power of β on either side of the equation.

$$\begin{aligned} \beta^0 : &\qquad H_0|a\rangle = E_a|a\rangle \\ \beta^1 : &\qquad H_0|b\rangle + \Delta|a\rangle = E_a|b\rangle + E_b|a\rangle \\ \beta^2 : &\qquad H_0|c\rangle + \Delta|b\rangle = E_a|c\rangle + E_b|b\rangle + E_c|a\rangle. \end{aligned} \tag{9.4}$$

The first equation simply states that E_a and $|a\rangle$ are an eigenvalue and eigenket of H_0. Physically, $|a\rangle$ is the state that we will find the system in if we slowly turn the knob back to zero after making a measurement of the energy. Henceforth we shall relabel E_a with E_0 and relabel $|a\rangle$ with $|E_0\rangle$, the zero reminding us of the association with $\beta = 0$ rather than implying that $|E_0\rangle$ is the ground state of the unperturbed system.

To determine E_b we multiply the second equation through by $\langle E_0|$:

$$\langle E_0|H_0|b\rangle + \langle E_0|\Delta|E_0\rangle = E_0\langle E_0|b\rangle + E_b. \tag{9.5}$$

Now from Table 2.1, $\langle E_0|H_0|b\rangle = (\langle b|H_0|E_0\rangle)^* = E_0\langle E_0|b\rangle$. Cancelling this with the identical term on the right, we are left with

$$E_b = \langle E_0|\Delta|E_0\rangle. \tag{9.6}$$

Thus *the first-order change in the energy is just the expectation value of the change in the Hamiltonian when the system is in its unperturbed state,* which makes good sense intuitively. This is the result that we anticipated in §§8.2.1 and 8.2.2 to estimate the effects on the allowed energies of hydrogen of the spin-orbit and hyperfine Hamiltonians.

To extract the second-order change in E we multiply the third of equations (9.4) by $\langle E_0|$. Cancelling $\langle E_0|H_0|c\rangle$ on $E_0\langle E_0|c\rangle$ by strict analogy with what we just did, we obtain

$$E_c = \langle E_0|\Delta|b\rangle - E_b\langle E_0|b\rangle. \tag{9.7}$$

To proceed further we have to determine $|b\rangle$, the first-order change in the state vector. Since the eigenkets $|E_n\rangle$ of H_0 form a complete set of states, we can write $|b\rangle$ as the sum

$$|b\rangle = \sum_k b_k|E_k\rangle. \tag{9.8}$$

In the second of equations (9.4) we replace $|b\rangle$ by this expansion and multiply through by $\langle E_m| \neq \langle E_0|$ to find

$$b_m = \frac{\langle E_m|\Delta|E_0\rangle}{E_0 - E_m}. \tag{9.9}$$

---

[1] Much interesting physics is associated with phenomena in which a small change in one variable can produce a large change in another (phase changes, narrow resonances, caustics, ...). In classical physics perturbation theory is bedevilled by such phenomena. In quantum mechanics this conjecture is more successful, but still untrustworthy as we shall see in §9.1.2.

This expression determines the coefficient of all kets in (9.8) that have energies that differ from the unperturbed value E_0. For the moment we assume that E_0 is a non-degenerate eigenvalue of H_0, so there is only one undetermined coefficient, namely that of $|E_0\rangle$. Fortunately we can argue that this coefficient can be taken to be zero from the requirement that $|E\rangle = |E_0\rangle + \beta|b\rangle + \mathrm{O}(\beta^2)$ remain correctly normalised. We can draw a useful analogy with changing a three-dimensional vector so that the condition $|\mathbf{r}| = 1$ is preserved; clearly we have to move $\mathbf{r}$ on the unit sphere and the first-order change in $\mathbf{r}$ is necessarily perpendicular to the original value of $\mathbf{r}$. The quantum-mechanical normalisation condition implies that as β increases $|E\rangle$ moves on a hypersphere in state space and $\langle b|E_0\rangle = 0$. So we exclude $|E_0\rangle$ from the sum in (9.8) and use this expression with the b_k given by (9.9) in (9.7) to find that the second-order change in E is

$$E_c = \sum_{k\neq 0} \frac{\langle E_0|\Delta|E_k\rangle\langle E_k|\Delta|E_0\rangle}{E_0 - E_k}. \tag{9.10}$$

9.1.1 Quadratic Stark effect

Let's apply the theory we've developed so far to a hydrogen atom that has been placed in an electric field $\mathbf{E} = -\nabla\Phi$. An externally imposed electric field is small compared to that inside an atom for field strengths up to $\mathcal{E} \simeq 5 \times 10^{11}\,\mathrm{V\,m^{-1}}$ (Problem 8.5) so perturbation theory should yield a good estimate of the shifts in energy level that ordinary fields effect. By the definition of the electrostatic potential Φ, the field changes the energy of the atom by

$$\delta E = e\{\Phi(\mathbf{x}_\mathrm{p}) - \Phi(\mathbf{x}_\mathrm{e})\}, \tag{9.11}$$

where $\mathbf{x}_\mathrm{p}$ and $\mathbf{x}_\mathrm{e}$ are the position vectors of the proton and electron, respectively. We assume that the field changes very little on the scale of the atom, and, as in §8.1, we define $\mathbf{r} \equiv \mathbf{x}_\mathrm{e} - \mathbf{x}_\mathrm{p}$. Then we may write

$$\delta E \simeq -e\mathbf{r}\cdot\nabla\Phi = e\mathbf{r}\cdot\mathbf{E}. \tag{9.12}$$

We orient our coordinate system so that $\mathbf{E}$ is parallel to the z axis and use the notation $\mathcal{E} = |\mathbf{E}|$. Then it is clear that the effect of imposing an external electric field is to add to the unperturbed Hamiltonian a term

$$\Delta = e\mathcal{E}z. \tag{9.13}$$

Suppose the atom is in its ground state $|100\rangle$, where the digits indicate the values of n, l and m. Then from equation (9.6) the first-order energy change in E is

$$E_b = e\mathcal{E}\langle 100|z|100\rangle. \tag{9.14}$$

In §4.1.4 we saw that the expectation value of any component of $\mathbf{x}$ vanishes in a state of well-defined parity. Since the ground-state ket $|100\rangle$ has well defined (even) parity, $E_b = 0$, and the change in E is dominated by the second-order term E_c. For our perturbation to the ground state of hydrogen, equation (9.10) becomes

$$E_c = e^2\mathcal{E}^2 \sum_{n=2}^{\infty} \sum_{\substack{l<n \\ |m|\leq l}} \frac{\langle 100|z|nlm\rangle\langle nlm|z|100\rangle}{E_1 - E_n}. \tag{9.15}$$

Symmetry considerations make it possible to simplify this sum dramatically. First, since $[L_z, z] = 0$ (Table 7.3), $z|nlm\rangle$ is an eigenfunction of L_z with eigenvalue m, and therefore orthogonal to $|100\rangle$ unless $m = 0$. Therefore in equation (9.15) only the terms with $m = 0$ contribute. Second, we can delete from the sum over l all even values of l because, as we saw in §4.1.4, the matrix elements of an odd-parity operator between states of the same parity vanish. In fact, a result proved in Problem 7.24 shows that the terms with $l = 1$ are the only non-vanishing terms in the sums over l in (9.15). Thus

$$E_c = e^2\mathcal{E}^2 \sum_{n=2}^{\infty} \frac{\langle 100|z|n10\rangle\langle n10|z|100\rangle}{E_1 - E_n}. \tag{9.16}$$

It is easy to understand physically why the change in E is proportional to $\mathcal{E}^2$. In response to the external electric field, the probability density of the atom's charge changes by an amount that is proportional to the coefficients b_k, and these coefficients are proportional to $\mathcal{E}$. That is, the field polarises the atom, generating a dipole moment $\mathbf{P}$ that is $\propto \mathcal{E}$. The dipole's energy is $-\mathbf{P}\cdot\mathbf{E}$, so the energy change caused by the field is proportional to $\mathcal{E}^2$.

9.1.2 Linear Stark effect and degenerate perturbation theory

Consider now the shift in the energy of the $n = 2$, $l = 0$ state of Hydrogen when an electric field is applied. The sum over k in (9.15) now includes the term

$$\frac{\langle 200|z|210\rangle\langle 210|z|200\rangle}{E_{20} - E_{21}},$$

which is infinite if we neglect the very small Lamb shift (§8.1.3), because the top is non-zero (Problem 8.11) and the difference of energies on the bottom vanishes. It hardly seems likely that a negligible field will produce an arbitrarily large change in the energy of the first excited state of hydrogen. So what did we do wrong?

Our error was to assume at the outset that a small stimulus produces a small response as we did when we wrote equations (9.2). Our infinite contribution to E_c can be traced to our expression (9.9) for b_m, which diverges as $E_m \to E_0$. That is, the change in the wavefunction that a given field produces is inversely proportional to the energy difference between the original state $|E_0\rangle$ and the state $|E_m\rangle$ we are pushing the system towards. This is an entirely reasonable result, analogous to what happens as we push a marble that lies at the bottom of a bowl: the distance the marble moves before coming into equilibrium depends on the curvature of the bowl. In the limit that the curvature goes to zero, and the bottom of the bowl becomes flat, an infinitesimal force will move the marble arbitrarily far, because all locations have the same energy. So we conclude that when the system's initial energy is a degenerate eigenvalue E_d of H_0, a tiny stimulus *is* liable to produce a big change in the *state* (but not the energy) of the system. Disaster will attend an attempt to calculate this abrupt change of state by the approach we have been developing.

So must we just give up in despair? No, because we can see that the only states that are going to acquire a non-negligible amplitude during the abrupt change are ones that have the same energy as E_d. That is, the state to which the system abruptly moves can be expressed as a linear combination of the kets belonging to E_d. In many cases of interest there are only a small number of these (four in the problem of hydrogen on which we are working). What we have to do is to diagonalise the matrix Δ_{ij} formed by Δ squeezed between all pairs of these kets. The eigenkets of Δ in this small subspace will be states of well-defined energy in the slightly perturbed system. As β is ramped up from zero to unity their energies will diverge from E_d. We conjecture that in the instant that β departs from zero, the system's state jumps to the eigenket with the lowest energy, and subsequently stays in this state as β increases. If this conjecture is correct, we should be able to use the perturbation theory we have developed provided we use as basis kets ones that diagonalise Δ as well as H_0.

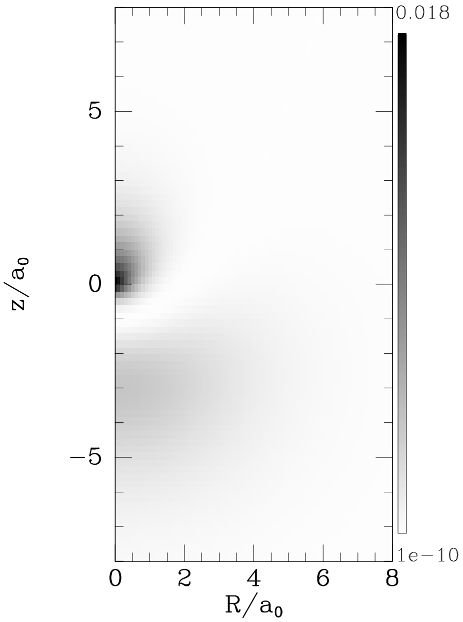


Figure 9.1 The charge distribution of the state $(|200\rangle + |210\rangle)/\sqrt{2}$ is axisymmetric. Here we plot the distribution in the (R, z) plane of cylindrical polar coordinates.

So let's diagonalise $e\mathcal{E}z$ in the 4-dimensional subspace of Hydrogen kets with $n = 2$. When we list the kets in the order $|200\rangle$, $|210\rangle$, $|211\rangle$, $|21{-}1\rangle$, the matrix of Δ looks like this

$$\Delta_{ij} = e\mathcal{E}\begin{pmatrix} 0 & a & 0 & 0 \\ a^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{where} \quad a = \langle 200|z|210\rangle. \tag{9.17}$$

From Problem 8.11 we have that $\langle 200|z|210\rangle = -3a_0$. It is now easy to show that the eigenvalues of $\mathbf{\Delta}$ are $\pm 3e\mathcal{E}a_0$ and 0, while appropriate eigenkets are $2^{-1/2}(1, \mp 1, 0, 0)$, $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$. We conclude that as soon as the slightest perturbation is switched on, the system is in the state of lowest energy, $|\psi\rangle = 2^{-1/2}\big(|200\rangle + |210\rangle\big)$, and we use this state to determine E_b. We find

$$\begin{aligned} E_b &= \tfrac{1}{2}e\mathcal{E}\big(\langle 200| + \langle 210|\big)z\big(|200\rangle + |210\rangle\big) \\ &= -3a_0 e\mathcal{E}. \end{aligned} \tag{9.18}$$

From our discussion of the quadratic Stark effect, we know that a change in E that is proportional to $\mathcal{E}$ requires the dipole moment $\mathbf{P}$ of an atom to be independent of $\mathcal{E}$. Since E_b is proportional to $\mathcal{E}$ we conclude that a hydrogen atom in the $n = 2$ state has a permanent electric dipole.

In classical physics this result is to be expected because the orbit of the electron would in general be elliptical, and the time-averaged charge density along the ellipse would be higher at the apocentre than at the pericentre,[2] because the electron lingers at the apocentre and rushes through the pericentre. Hence the centre of charge would lie on the opposite side of the geometrical centre of the ellipse from the focus, where the proton's cancelling charge lies. Thus, if the electron's orbit were a perfect Kepler ellipse, the atom would have a permanent electric dipole moment parallel to the orbit's major axis. Any deviation of the radial force field from $F \propto r^{-2}$ will cause the major axis of the ellipse to precess, and therefore the time-averaged polarisation of the atom to be zero. In hydrogen the force-field deviates verify little from an inverse-square law, so the precession occurs very slowly in the classical picture. Consequently, even a weak external field can prevent precession and thus give rise to a steady electric dipole.

---

[2] An orbit's **apocentre** is the point furthest from the attracting body, while the **pericentre** is the point nearest that body.

In the quantum-mechanical picture, shielding shifts the energy of the S state below that of the P states, thus ensuring that, in the absence of an imposed field, the atom is spherical and has no dipole moment. An electric field deprives L^2 of its status as a constant of motion because the field can apply a torque to the atom. Shielding is a very weak effect in hydrogen (because it relies on the vacuum's virtual electrons and positrons), so the S state lies very little below the P states and in even a weak electric field this offset becomes irrelevant. The lowest-energy state becomes $(|200\rangle + |210\rangle)/\sqrt{2}$. This is not an eigenket of L^2 but it is an eigenket of L_z with eigenvalue zero. Thus its angular momentum is perpendicular to the field, as we expect from the classical picture of a Kepler ellipse with its major axis parallel to $\mathbf{E}$. Figure 9.1 shows that in this state the charge distribution comprises a dense cloud around the origin and an extended cloud centred on $R = 0$, $z \simeq -3a_0$. We can think of these clouds as arising from pericentre and apocentre, respectively, of eccentric orbits that have their major axes roughly aligned with the negative z axis. The integral $\int \mathrm{d}^3\mathbf{x}, z|\psi|^2 = -3a_0$, so in this state the atom has dipole moment $P = +3ea_0$.

9.1.3 Effect of an external magnetic field

When an atom is placed in a magnetic field, the wavelengths of lines in its spectrum change slightly. Much of quantum mechanics emerged from attempts to understand this phenomenon. We now use perturbation theory to explain it.

In §3.3 we discussed the motion of a free particle in a uniform magnetic field. Our starting point was the Hamiltonian (3.30), which governs the motion of a free particle of mass m and charge Q in the magnetic field produced by the vector potential $\mathbf{A}$. This is the Hamiltonian of a free particle, $p^2/2m$, with $\mathbf{p}$ replaced by $\mathbf{p} - Q\mathbf{A}$. Hence we can incorporate the effects of a magnetic field on a hydrogen atom by replacing $\mathbf{p}_\mathrm{n}$ and $\mathbf{p}_\mathrm{e}$ in the gross-structure Hamiltonian (8.1) with $\mathbf{p}_\mathrm{p} - e\mathbf{A}$ and $\mathbf{p}_\mathrm{e} + e\mathbf{A}$, respectively. With $Z = 1$ the kinetic energy term in the Hamiltonian then becomes

$$\begin{aligned} H_\mathrm{KE} &\equiv \frac{(\mathbf{p}_\mathrm{p} - e\mathbf{A})^2}{2m_\mathrm{p}} + \frac{(\mathbf{p}_\mathrm{e} + e\mathbf{A})^2}{2m_\mathrm{e}} \\ &= \frac{p_\mathrm{p}^2}{2m_\mathrm{p}} + \frac{p_\mathrm{e}^2}{2m_\mathrm{e}} + \tfrac{1}{2}e\left\{\left(\frac{\mathbf{p}_\mathrm{e}}{m_\mathrm{e}} - \frac{\mathbf{p}_\mathrm{p}}{m_\mathrm{p}}\right)\cdot\mathbf{A} + \mathbf{A}\cdot\left(\frac{\mathbf{p}_\mathrm{e}}{m_\mathrm{e}} - \frac{\mathbf{p}_\mathrm{p}}{m_\mathrm{p}}\right)\right\} + \mathrm{O}(A^2) \end{aligned} \tag{9.19}$$

We neglect the terms that are $\mathrm{O}(A^2)$ on the grounds that when the field is weak enough for the $\mathrm{O}(A)$ terms to be small compared to the terms in the gross-structure Hamiltonian, the $\mathrm{O}(A^2)$ terms are negligible.

Equation (8.4) and the corresponding equation for $\partial/\partial\mathbf{x}_\mathrm{p}$ imply that

$$\mathbf{p}_\mathrm{e} = \frac{m_\mathrm{e}}{m_\mathrm{e} + m_\mathrm{p}}\mathbf{p}_\mathbf{X} + \mathbf{p}_\mathbf{r} \quad \text{and} \quad \mathbf{p}_\mathrm{p} = \frac{m_\mathrm{p}}{m_\mathrm{e} + m_\mathrm{p}}\mathbf{p}_\mathbf{X} - \mathbf{p}_\mathbf{r}, \tag{9.20}$$

where $\mathbf{p}_\mathbf{X}$ is the momentum associated with the centre of mass coordinate $\mathbf{X}$, while $\mathbf{p}_\mathbf{r}$ is the momentum of the reduced particle. From the algebra that leads to equation (8.6a) we know that the first two terms on the right of the second line of equation (9.19) reduce to the kinetic energy of the centre-of-mass motion and of the reduced particle. Using equations (9.20) in the remaining terms on the right of equation (9.19) yields

$$H_\mathrm{KE} = \frac{p_\mathbf{X}^2}{2(m_\mathrm{e} + m_\mathrm{p})} + \frac{p_\mathbf{r}^2}{2\mu} + \frac{e}{2\mu}(\mathbf{p}_\mathbf{r}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{p}_\mathbf{r}), \tag{9.21}$$

where μ is the mass of the reduced particle (eq. 8.6b). It follows that an external magnetic field adds to the gross-structure Hamiltonian of a hydrogen atom a perturbing Hamiltonian

$$H_B = \frac{e}{2\mu}(\mathbf{p}_\mathbf{r}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{p}_\mathbf{r}). \tag{9.22}$$

On the scale of the atom the field is likely to be effectively homogeneous, so we may take $\mathbf{A} = \frac{1}{2}\mathbf{B} \times \mathbf{r}$ (page 48). Then H_B becomes

$$H_B = \frac{e}{4m_\text{e}}(\mathbf{p} \cdot \mathbf{B} \times \mathbf{r} + \mathbf{B} \times \mathbf{r} \cdot \mathbf{p}), \tag{9.23}$$

where we have approximated μ by m_e and dropped the subscript on $\mathbf{p}$. The two terms in the bracket on the right can both be transformed into $\mathbf{B} \cdot \mathbf{r} \times \mathbf{p} = \hbar\mathbf{B} \cdot \mathbf{L}$ because (i) these scalar triple products involve only products of *different* components of the three vectors, and (ii) $[x_i, p_j] = 0$ for $i \neq j$. Hence, we do not need to worry about the order of the $\mathbf{r}$ and $\mathbf{p}$ operators and can exploit the usual invariance of a scalar triple product under cyclic interchange of its vectors.

If an atom has more than one unpaired ('valence') electron, each electron will contribute a term of this form to the overall Hamiltonian. We can fold these separate contributions into a single contribution H_B by interpreting $\mathbf{L}$ as the sum of the angular-momentum operators of the individual electrons.

In §8.2.1 we discussed terms that must be added to hydrogen's gross-structure Hamiltonian to account for the effects of the electron's intrinsic dipole moment. We found that the coupling with an external field is generated by the Zeeman spin Hamiltonian (8.68). Adding this to the value of H_B that we have just computed, and orienting our coordinate system so that the z axis is parallel to $\mathbf{B}$, we arrive at our final result, namely that a uniform magnetic field introduces a perturbation

$$H_{Bs} = \frac{e\hbar}{2m_\text{e}}B(L_z + 2S_z) = \mu_\text{B}B(J_z + S_z), \tag{9.24}$$

where $\mathbf{S}$ is the sum of the spin operators of all the valence electrons.

The Hamiltonian formed by adding H_{Bs} to the gross-structure Hamiltonian (8.1) commutes with L^2, L_z, S^2 and S_z. Its eigenkets are simply the eigenkets of the gross-structure Hamiltonian upgraded to include eigenvalues of S^2 and S_z. The only difference from the situation we studied in §8.1 is that the energies of these eigenkets now depend on both L_z and S_z. Hence, each energy level of the gross-structure Hamiltonian is split by the magnetic field into as many sub-levels as $m_l + 2m_s$ can take. For example, if $l = 0$ and $s = \frac{1}{2}$, there are two sublevels, while when $l = 1$ and $s = \frac{1}{2}$ there are five levels in which $m_l + 2m_s$ ranges between ± 2.

In practice the perturbation H_{Bs} always acts in conjunction with the spin-orbit perturbation H_SO of equation (8.69).[3] The general case in which H_{Bs} and H_SO are comparable, requires numerical solution. The extreme cases in which one operator is larger than the other can be handled analytically.

Paschen–Back effect In a sufficiently strong magnetic field, H_SO affects the atom much less than H_{Bs}, so H_SO simply perturbs the eigenkets of the Hamiltonian formed adding H_{Bs} to the gross-structure Hamiltonian. The change in the energy of the state $|n, l, m_l, s, m_s\rangle$ is

$$\begin{aligned} E_b &= \langle n, l, m_l, s, m_s|H_\text{SO}|n, l, m_l, s, m_s\rangle \\ &= \zeta\langle n, l, m_l, s, m_s|\mathbf{L} \cdot \mathbf{S}|n, l, m_l, s, m_s\rangle, \end{aligned} \tag{9.25}$$

where ζ is a number with dimensions of energy that is independent of m_l and m_s. By writing $\mathbf{L} \cdot \mathbf{S} = \frac{1}{2}(L_+S_- + L_-S_+) + L_zS_z$ (eq. 7.137) we see that $\langle \mathbf{L} \cdot \mathbf{S}\rangle = m_l m_s$. So in a strong magnetic field the eigenenergies are

$$E_\text{gross} + \mu_\text{B}B(m_l + 2m_s) + \zeta m_l m_s. \tag{9.26}$$

---

[3] There is no spin-orbit coupling for an S state, but an allowed spectral line from an S state will connect to a P state for which there is spin-orbit coupling. Hence the frequencies of allowed transitions inevitably involve spin-orbit coupling.

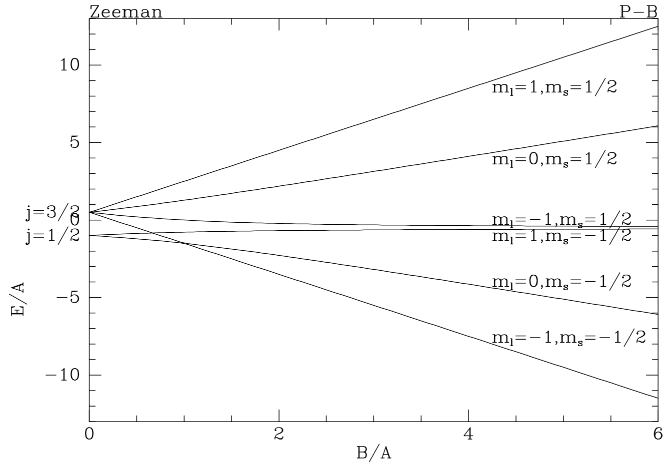


Figure 9.2 Eigenvalues of the spin-dependent Hamiltonian $A\mathbf{L}\cdot\mathbf{S} + B(L_z + 2S_z)$ as functions of B/A for the case $l = 1$, $s = \frac{1}{2}$. The right side of the diagram (field strong compared to spin-orbit coupling) quantifies quantifies the Paschen–Back effect, while the left side of the diagram quantifies the Zeeman effect (weak field). The top and bottom lines on the extreme right show the energies of the states $|1, 1\rangle|+\rangle$ and $|1, -1\rangle|-\rangle$, which are eigenstates of the full Hamiltonian for all values of B/A.

The levels on the extreme right of Figure 9.2 show the energies described by this formula in the case that $l = 1$ and $s = \frac{1}{2}$. The fact that in a strong magnetic field an atom's energies depend on m_l and m_s in this way is known as the **Paschen–Back effect**.

Zeeman effect In a sufficiently weak magnetic field, $H_{\rm SO}$ affects the atom more strongly than H_{Bs}. Then spin-orbit coupling assigns different energies to states that differ in j. Consequently, when we use perturbation theory to calculate the smaller effect of an imposed magnetic field, the degenerate eigenspace in which we have to work is that spanned by the states that have given values of j, l and s but differ in their eigenvalues m of J_z. Fortunately, H_{Bs} is already diagonal within this space because $[J_z, S_z] = 0$. So the shift in the energy of each state is simply

$$E_b = \langle j, m, l, s|H_{Bs}|j, m, l, s\rangle = \mu_{\rm B}\big(m + \langle j, m, l, s|S_z|j, m, l, s\rangle\big). \tag{9.27}$$

As we saw in §7.5, our basis states do not have well-defined values of S_z – in general they are linear combinations of eigenstates of L_z and S_z:

$$|j, m, l, s\rangle = \sum_{m'=-s}^{s} c_{m'}|l, m - m'\rangle|s, m'\rangle, \tag{9.28}$$

where the coefficients $c_{m'}$ are Clebsch–Gordan coefficients (eq. 7.146). In any concrete case it is straightforward to calculate the required expectation value of S_z from this expansion. However, a different approach yields a general formula that was important historically.

In the classical picture, spin-orbit coupling causes the vector $\mathbf{S}$ to precess around the invariant vector $\mathbf{J}$. Hence, in this picture the expectation value of $\mathbf{S}$ is equal to the projection of $\mathbf{S}$ onto $\mathbf{J}$.[4] The classical vector triple product formula enables us to express $\mathbf{S}$ in terms of this projection:

$$\mathbf{J}\times(\mathbf{S}\times\mathbf{J}) = J^2\mathbf{S} - (\mathbf{S}\cdot\mathbf{J})\mathbf{J} \quad \text{so} \quad J^2\mathbf{S} = (\mathbf{S}\cdot\mathbf{J})\mathbf{J} + \mathbf{J}\times(\mathbf{S}\times\mathbf{L}). \tag{9.29}$$

In the classical picture, the expectation value of the vector triple product on the right side vanishes. If its quantum expectation value were to vanish,

---

[4] This heuristic argument is often referred to as the **vector model**.

the expectation value of the z component of the equation would relate $\langle S_z\rangle$, which we require, to the expectation values of operators that have the states $|j,m,l,s\rangle$ as eigenstates, so our problem would be solved. Motivated by these classical considerations, let's investigate the operator

$$\mathbf{G} \equiv \mathbf{J}\times(\mathbf{S}\times\mathbf{L}) \quad \text{so} \quad G_i \equiv \sum_{jklm} \epsilon_{ijk} J_j \epsilon_{klm} S_l L_m. \tag{9.30}$$

It is straightforward to check that its components commute with the angular-momentum operators J_i in the way we expect the components of a vector to do:

$$[J_i, G_j] = \mathrm{i}\sum_k \epsilon_{ijk} G_k. \tag{9.31}$$

From equation (9.30) it is also evident that $\mathbf{J}\cdot\mathbf{G} = 0$. In Problem 7.24 identical conditions on the operators $\mathbf{L}$ and $\mathbf{x}$ suffice to prove that $\langle x\rangle = 0$ in any state that is an eigenket of L^2. So the steps of that proof can be retraced with $\mathbf{L}$ replaced by $\mathbf{J}$ and $\mathbf{x}$ replaced by $\mathbf{G}$ to show that for the states of interest $\langle\mathbf{G}\rangle = 0$.

Now that we have established that the quantum-mechanical expectation value of $\mathbf{G}$ does indeed vanish, we reinterpret equation (9.29) as an operator equation, and, from the expectation value of its z component, deduce

$$\langle j,m,l,s|S_z|j,m,l,s\rangle = \frac{\langle\mathbf{J}\cdot\mathbf{S}\rangle\, m}{j(j+1)}. \tag{9.32}$$

From equation (8.70) we have

$$\mathbf{J}\cdot\mathbf{S} = \mathbf{L}\cdot\mathbf{S} + S^2 = \tfrac{1}{2}(J^2 - L^2 + S^2), \tag{9.33}$$

so we find

$$E_B = m g_{\mathrm{L}} \mu_{\mathrm{B}} B \text{ where } g_{\mathrm{L}} \equiv \left(1 + \frac{j(j+1) - l(l+1) + s(s+1)}{2j(j+1)}\right). \tag{9.34}$$

The factor g_{L} is called the **Landé g factor**. In the early days of quantum theory, when the Bohr atom was taken seriously, people expected the magnetic moment of an electron to be $\pm\mu_{\mathrm{B}}$ and therefore thought a magnetic field would shift energy levels by $\pm\mu_{\mathrm{B}}B$. Equation (9.34) states that the actual shift is $m g_{\mathrm{L}}$ times this. When this factor differed from unity, they spoke of an **anomalous Zeeman effect**.

The left hand side of Figure 9.2 shows the energy levels described by equation (9.34) in the case $l = 1$, $s = \frac{1}{2}$. The possible values of j are $\frac{3}{2}$ and $\frac{1}{2}$, and the magnetic field splits each of these spin-orbit levels into $2j+1$ components.

9.2 Variational principle

We now describe a method of estimating energy levels, especially a system's ground-state energy, that does not involve breaking the Hamiltonian down into a part that has known eigenkets and an additional perturbation. In Chapter 10 we shall show that this method yields quite an accurate value for the ionisation energy of helium.

Let H be the Hamiltonian for which we require the eigenvalues E_n and the associated eigenkets $|n\rangle$. We imagine expanding an arbitrary state $|\psi\rangle = \sum_n a_n|n\rangle$ as a linear combination of these eigenkets, and then calculate the expectation value of H in this state as

$$\langle H\rangle = \langle\psi|H|\psi\rangle = \frac{\sum_i |a_i|^2 E_i}{\sum_j |a_j|^2}, \tag{9.35}$$

where we have included the sum of the $|a_j|^2$ on the bottom to cover the possibility that $|\psi\rangle$ is not properly normalised. $\langle H\rangle$ is manifestly independent of the phase of a_i. We investigate the stationary points of $\langle H\rangle$ with respect to the moduli $|a_i|$ by differentiating equation (9.35) with respect to them:

$$\frac{\partial\langle H\rangle}{\partial|a_k|}=\frac{2|a_k|E_k}{\sum_j|a_j|^2}-\frac{2|a_k|\sum_i|a_i|^2E_i}{\left(\sum_j|a_j|^2\right)^2}. \tag{9.36}$$

Equating this derivative to zero, we find that the conditions for a stationary point of $\langle H\rangle$ are

$$0=|a_k|\left(E_k\sum_i|a_i|^2-\sum_i|a_i|^2E_i\right)\quad(k=0,1,\ldots) \tag{9.37}$$

These equations are trivially solved by setting $a_k=0$ for every k, but then $|\psi\rangle=0$ so the solution is of no interest. For any value of k for which $a_k\neq0$, we must have

$$E_k=\frac{\sum_i|a_i|^2E_i}{\sum_i|a_i|^2}. \tag{9.38}$$

Since the right side of this equation does not depend on k, the equation can be satisfied for at most one value of k, and it clearly is satisfied if we set $a_i=0$ for $i\neq k$ and $a_k=1$, so $|\psi\rangle=|k\rangle$. This completes the proof of **Rayleigh's theorem**: *The stationary points of the expectation value of an Hermitian operator occur at the eigenstates at that operator. Moreover, all eigenstates provide stationary points of the operator.* That is, for general $|\psi\rangle$ the number of $\langle\psi|H|\psi\rangle$ isn't equal to an eigenvalue of H, but if $\langle\psi|H|\psi\rangle$ is stationary with respect to $|\psi\rangle$ in the sense that it doesn't change when $|\psi\rangle$ is changed by a small amount, Rayleigh's theorem tells us that the number $\langle\psi|H|\psi\rangle$ *is* an eigenvalue of H. Problem 9.9 gives a geometrical interpretation of Rayleigh's theorem.

The stationary point associated with the ground state is a minimum of $\langle H\rangle$. To see that this is so, we subtract the ground-state energy E_0 from both sides of equation (9.35) and have

$$\langle H\rangle-E_0=\frac{\sum_i|a_i|^2(E_i-E_0)}{\sum_j|a_j|^2}. \tag{9.39}$$

Both the top and bottom of the fraction on the right are non-negative, so $\langle H\rangle\geq E_0$. The stationary points of $\langle H\rangle$ associated with excited states are saddle points (Problem 9.12).

The practical use of Rayleigh's theorem is this. We write down a trial wavefunction $\psi_{\mathbf{a}}(\mathbf{x})$ that depends on a number of parameters $a_1,\ldots,a_N$. These might, for example, be the coefficients in an expansion of $\psi_{\mathbf{a}}$ as a linear combination of some convenient basis functions $u_i(\mathbf{x})$

$$\psi_{\mathbf{a}}(\mathbf{x})\equiv\sum_{i=1}^N a_iu_i(\mathbf{x}). \tag{9.40}$$

More often the a_i are parameters in a functional form that is motivated by some physical argument. For example, in Chapter 10 we will treat the variable Z that appears in the hydrogenic wavefunctions of §8.1.2 as one of the a_i. Then we use $\psi_{\mathbf{a}}$ to calculate $\langle H\rangle$ as a function of the a_i and find the stationary points of this function. The minimum value of $\langle H\rangle$ that we obtain in this way clearly provides an upper limit on the ground-state energy E_0. Moreover, since $\langle H\rangle$ is stationary for the ground-state wavefunction, $\langle H\rangle-E_0$ increases only quadratically in the difference between $\psi_{\mathbf{a}}$ and the ground-state wavefunction ψ_0. Hence, with even a mediocre fit to ψ_0 this upper limit will lie close to E_0. This approach to finding eigenvalues and eigenfunctions of the Hamiltonian is called the **variational principle**.

In Problems 9.10 and 9.11 you can explore how the variational principle works in a simple case.

9.3 Time-dependent perturbation theory

We now describe a way of obtaining approximate solutions to the TDSE (2.26) that we shall use to study both scattering of particles and the emission and absorption of radiation by atoms and molecules.

Consider the evolution of a system that is initially in a state that is nearly, but not quite, in a stationary state. Specifically, at $t = 0$ it is in the Nth eigenstate of a Hamiltonian H_0 that differs by only a small, possibly time-dependent, operator V from the true Hamiltonian H:

$$H = H_0 + V. \tag{9.41}$$

Inspired by (2.32) we expand the solution to the TDSE for this H in the form

$$|\psi\rangle = \sum_n a_n(t)\mathrm{e}^{-\mathrm{i}E_n t/\hbar}|E_n\rangle, \tag{9.42}$$

where $|E_n\rangle$ is a (time-independent) eigenket of H_0 with eigenvalue E_n. This expansion doesn't restrict $|\psi\rangle$ because $\{|E_n\rangle\}$ is a complete set and the functions $a_n(t)$ are arbitrary. Substituting it into the TDSE we have

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial t} &= (H_0 + V)|\psi\rangle = \sum_n \Big(E_n|E_n\rangle + V|E_n\rangle\Big)a_n\mathrm{e}^{-\mathrm{i}E_n t/\hbar} \\ &= \sum_n \big(\mathrm{i}\hbar\dot{a}_n + E_n a_n\big)\mathrm{e}^{-\mathrm{i}E_n t/\hbar}|E_n\rangle. \end{aligned} \tag{9.43}$$

We simplify this by multiplying through by $\langle E_k|$:

$$\mathrm{i}\hbar\dot{a}_k\mathrm{e}^{-\mathrm{i}E_k t/\hbar} = \sum_n a_n\mathrm{e}^{-\mathrm{i}E_n t/\hbar}\langle E_k|V|E_n\rangle. \tag{9.44}$$

This constitutes a set of linear ordinary differential equations for the $a_n(t)$ which must be solved subject to the boundary conditions $a_N(0) = 1$ and $a_n(0) = 0$ for $n \neq N$. Hence, at the earliest times the term on the right of (9.44) with $n = N$ will dominate the equation of motion of a_k with $k \neq N$, and we have the approximation

$$\dot{a}_k \simeq -\frac{\mathrm{i}}{\hbar}\mathrm{e}^{-\mathrm{i}(E_N - E_k)t/\hbar}\langle E_k|V|E_N\rangle. \tag{9.45}$$

We now assume that any time dependence of V takes the form $V(t) = V_0\mathrm{e}^{\mathrm{i}\omega t}$, where V_0 is a time-independent operator. This assumption is in practice not very restrictive because the theory of Fourier analysis enables us to express any operator of the form $V_0 f(t)$, where f is an arbitrary function, as a linear combination of sinusoidally varying operators. Replacing V by $V_0\mathrm{e}^{\mathrm{i}\omega t}$ in equation (9.45) and integrating from $t = 0$ we find

$$a_k(t) = \frac{\langle E_k|V_0|E_N\rangle}{E_N - E_k - \hbar\omega}\left[\mathrm{e}^{-\mathrm{i}(E_N - E_k - \hbar\omega)t'/\hbar}\right]_0^t, \tag{9.46}$$

so the probability that after t the system has made the transition to the kth eigenstate of H_0 is

$$\begin{aligned} P_k(t) &= |a_k|^2 \\ &= \frac{|\langle E_k|V_0|E_N\rangle|^2}{(E_N - E_k - \hbar\omega)^2}\left\{2 - 2\cos\left(\frac{(E_N - E_k - \hbar\omega)t}{\hbar}\right)\right\} \\ &= 4|\langle E_k|V_0|E_N\rangle|^2\frac{\sin^2\left((E_N - E_k - \hbar\omega)t/2\hbar\right)}{(E_N - E_k - \hbar\omega)^2}. \end{aligned} \tag{9.47}$$

For a time of order $\hbar/(E_N - E_k - \hbar\omega)$ this expression grows like t^2. Subsequently it oscillates.

9.3.1 Fermi golden rule

Consider now a case in which V is a time-independent perturbation, so $\omega = 0$. In applications of interest H_0 has a large number of eigenvalues E_k within an interval of width h/t of E, and we are typically interested in the probability that the system has made the transition to *any* one of these states. Hence we sum the P_k over k. Let there be $g(E)\,\mathrm{d}E$ eigenvalues in the interval $(E+\mathrm{d}E, E)$. Then the total transition probability is

$$\begin{aligned}\sum_k P_k(t) &= 4\int \mathrm{d}E\, g(E)\, |\langle E|V|E_N\rangle|^2\, \frac{\sin^2\left((E_N-E)t/2\hbar\right)}{(E_N-E)^2} \\ &= \frac{2}{\hbar}\int \mathrm{d}x\, g(E_N-2\hbar x)\, |\langle E_N-2\hbar x|V|E_N\rangle|^2\, \frac{\sin^2(xt)}{x^2},\end{aligned} \tag{9.48}$$

where we've introduced a new variable, $x = (E_N - E)/2\hbar$. For given t, the function $f_t(x) \equiv \sin^2(xt)/x^2$ is dominated by a bump around the origin that is of height t^2 and width $2\pi/t$. Hence, the area under the bump is proportional to t and in the limit of large t,

$$\frac{\sin^2(xt)}{x^2} \propto t\delta(x). \tag{9.49}$$

We find the constant of proportionality by differentiating $\int \mathrm{d}x\, f_t$ with respect to t:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty} \mathrm{d}x\, \frac{\sin^2(xt)}{x^2} = \int_{-\infty}^{\infty} \mathrm{d}x\, \frac{\sin(2xt)}{x} = \pi. \tag{9.50}$$

Hence

$$\lim_{t\to\infty} \frac{\sin^2(xt)}{x^2} = \pi t\delta(x). \tag{9.51}$$

Inserting this relation in (9.48) and integrating over x, we have finally

$$\sum_k P_k = \frac{2\pi t}{\hbar} g(E_N)\, |\langle \mathrm{out}|V|\mathrm{in}\rangle|^2. \tag{9.52}$$

This simple result is **Fermi's golden rule**[5] of perturbation theory. The coefficient of t on the right gives the rate at which the system leaks out of the state $|\mathrm{in}\rangle$ to other states of the same energy.

9.3.2 Radiative transition rates

We now use equation (9.47) to calculate the rate at which an electromagnetic wave induces an atom to make radiative transitions between its stationary states. Our treatment is valid when the quantum uncertainty in the electromagnetic field may be neglected, and the field treated as a classical object. This condition is satisfied, for example, in a laser, or at the focus of the antenna of a radio telescope.

Whereas in our derivation of Fermi's golden rule, the system had some density of states at the energy of the initial state and we summed the transition probabilities by integrating over the energy of the final state, now we argue that the electromagnetic field has non-negligible power over a continuum of frequencies, and we integrate over frequency. The quantity $|\langle E_k|V_0|E_N\rangle|^2$ that occurs in equation (9.47) is now a function of frequency ω. We argue that each single-frequency contribution to the electromagnetic field induces transitions independently contributions at other frequencies. Hence equation (9.47) is valid if we make the substitution

$$|\langle E_k|V_0|E_N\rangle|^2 \quad\to\quad \mathrm{d}\omega\, \frac{\mathrm{d}}{\mathrm{d}\omega}|\langle E_k|V_0|E_N\rangle|^2, \tag{9.53}$$

---

[5] The golden rule was actually first given by Dirac, P.A.M., 1927, Proc. Roy. Soc. A, **114**, 243

which isolates the contribution to $|\langle E_k|V_0|E_N\rangle|^2$ from a frequency interval of infinitesimal width. Once we have evaluated the transition probability $P_k(t)$ that is generated by this frequency interval, we set t to a large value and sum over all intervals by integrating with respect to ω. That is we evaluate

$$\int \mathrm{d}\omega\, P_k = \frac{1}{\hbar^2}\int \mathrm{d}\omega\, \frac{\sin^2(xt)}{x^2}\frac{\mathrm{d}}{\mathrm{d}\omega}|\langle E_k|V_0|E_N\rangle|^2, \tag{9.54}$$

where $x \equiv (\hbar\omega + E_k - E_N)/2\hbar$. We change the integration variable to x using $\mathrm{d}x = \frac{1}{2}\mathrm{d}\omega$ and exploit equation (9.51) to evaluate the integral. The result is

$$\int \mathrm{d}\omega\, P_k = \frac{2\pi t}{\hbar^2}\frac{\mathrm{d}}{\mathrm{d}\omega}|\langle E_k|V_0|E_N\rangle|^2, \quad \text{with} \quad \omega = (E_N - E_k)/\hbar. \tag{9.55}$$

The coefficient of t on the right is the rate at which the sinusoidally oscillating perturbation causes transitions from $|E_N\rangle$ to $|E_k\rangle$. This rate vanishes unless there is a component of the perturbation that oscillates at the angular frequency $(E_N - E_k)/\hbar$ that we associate with the energy difference between the initial and final states. This makes perfect sense physically because we know from §3.2.1 that when there is quantum uncertainty as to whether the system is in two states that differ in energy by ΔE, the system oscillates physically with angular frequency $\Delta E/\hbar$. If the system bears electromagnetic charge, these oscillations are liable to transfer energy either into or out of the oscillations of the ambient electromagnetic field at this frequency.

To proceed further we need an expression for the derivative of the matrix element in equation (9.55). In vacuo the electric field of an electromagnetic wave is divergence free, being entirely generated by Faraday's law, $\nabla\times\mathbf{E} = -\partial\mathbf{B}/\partial t$. It follows that the whole electromagnetic field of the wave can be described by the vector potential $\mathbf{A}$ through the equations $\mathbf{B} = \nabla\times\mathbf{A}$ and $\mathbf{E} = -\partial\mathbf{A}/\partial t$. We assume that we are dealing with a plane wave

$$\mathbf{A} = \mathbf{A}_0\cos(\mathbf{k}\cdot\mathbf{x} - \omega t), \tag{9.56}$$

where $\mathbf{A}_0$ is a constant vector and $\mathbf{k}$ is the wavevector. Electromagnetic waves are transverse in the sense that $\mathbf{E}$ and $\mathbf{B}$ are perpendicular to the direction of propagation, in this case $\mathbf{k}$. From equation (9.56) we have

$$\mathbf{E} = -\omega\mathbf{A}_0\cos(\mathbf{k}\cdot\mathbf{x} - \omega t), \tag{9.57}$$

so $\mathbf{E}$ is parallel to $\mathbf{A}_0$ and the equation

$$\mathbf{k}\cdot\mathbf{A}_0 = 0 \tag{9.58}$$

must hold.

In §9.1.3 we saw that an external electromagnetic field adds to an atom's Hamiltonian the perturbing term (9.22) for each electron. In the present case the perturbation is

$$V = \frac{e}{2m_\mathrm{e}}\{\mathbf{p}\cdot\mathbf{A}_0\cos(\mathbf{k}\cdot\mathbf{x} - \omega t) + \cos(\mathbf{k}\cdot\mathbf{x} - \omega t)\mathbf{A}_0\cdot\mathbf{p}\}. \tag{9.59}$$

By virtue of equation (9.58), $\mathbf{A}_0\cdot\mathbf{p}$ commutes with $\mathbf{k}\cdot\mathbf{x}$ because a component of momentum always commutes with a perpendicular component of position. Since $\mathbf{A}_0$ is a constant, it commutes with $\mathbf{p}$. So we can simplify V to

$$V = \frac{e}{m_\mathrm{e}}\mathbf{A}_0\cdot\mathbf{p}\,\cos(\mathbf{k}\cdot\mathbf{x} - \omega t) = \frac{e}{2m_\mathrm{e}}\mathbf{A}_0\cdot\mathbf{p}\left(\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x}-\omega t)} + \mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\mathbf{x}-\omega t)}\right). \tag{9.60}$$

We now make the approximation that the electromagnetic wavelength is much bigger than the characteristic size of the atom or molecule. This is a good one providing the atom or molecule moves between states that are separated in energy by much less than $\alpha m_{\rm e}c^2$ (Problem 9.15), as will be the case for waves with frequencies that are less than those of soft X-rays. In this case we will have $\mathbf{k}\cdot\mathbf{x} \ll 1$ for all locations $\mathbf{x}$ in the atom or molecule at which there is significant probability of finding an electron. When this condition is satisfied, it makes sense to expand the exponentials in equation (9.60) as power series and discard all but the constant term. We then have

$$V = \frac{e}{2m_{\rm e}}\,\mathbf{A}_0\cdot\mathbf{p}\left(\mathrm{e}^{-\mathrm{i}\omega t} + \mathrm{e}^{\mathrm{i}\omega t}\right), \tag{9.61}$$

where we have retained the exponentials in time because large values of t cannot be excluded in the way that we can exclude large values of $\mathbf{x}$. Finally, we note that in the gross-structure Hamiltonian H_0, $\mathbf{p}$ occurs only in the term $p^2/2m_{\rm e}$, so $[H_0,\mathbf{x}] = -\mathrm{i}(\hbar/m_{\rm e})\mathbf{p}$. When we use this relation to eliminate $\mathbf{p}$ from equation (9.61), we have finally

$$V = \mathrm{i}\frac{eA_0}{2\hbar}(\mathrm{e}^{-\mathrm{i}\omega t} + \mathrm{e}^{\mathrm{i}\omega t})[H_0, z], \tag{9.62}$$

where we have chosen to make the z axis parallel to $\mathbf{A}_0$. Thus a plane electromagnetic wave gives rise to perturbations with both positive and negative frequencies. Above we derived the frequency condition $\omega = (E_N - E_k)/\hbar$ for transitions from $|E_N\rangle$ to $|E_k\rangle$, so the negative frequency perturbation is associated with excitation of the system ($E_k > E_N$), while the positive frequency perturbation is associated with radiative decays.

Recalling that the quantity V_0 that appears in equation (9.55) is defined by $V = V_0\mathrm{e}^{\mathrm{i}\omega t}$, and using equation (9.62) to eliminate V_0 from equation (9.55), we find that the transition rate is

$$\begin{aligned} R \equiv \frac{\mathrm{d}}{\mathrm{d}t}\int \mathrm{d}\omega\, P_k &= \frac{\pi e^2}{2\hbar^4}\big|\langle E_k|[H_0,z]|E_N\rangle\big|^2 \frac{\mathrm{d}A_0^2}{\mathrm{d}\omega}\bigg|_{\omega=(E_N-E_k)/\hbar} \\ &= \frac{\pi e^2}{2\hbar^4}(E_k - E_N)^2\big|\langle E_k|z|E_N\rangle\big|^2 \frac{\mathrm{d}A_0^2}{\mathrm{d}\omega}\bigg|_{\omega=(E_N-E_k)/\hbar}. \end{aligned} \tag{9.63}$$

We can relate A_0^2 to the energy density of the electromagnetic field,

$$\rho = \frac{1}{2\mu_0}\{(E/c)^2 + B^2\}. \tag{9.64}$$

In an electromagnetic wave, the electric and magnetic energy densities are equal, so the overall energy density is twice the contribution given above from the electric field. Moreover, $E^2 = (\partial A/\partial t)^2 = \omega^2 A_0^2\cos^2(\mathbf{k}\cdot\mathbf{x} - \omega t)$, so the time-averaged energy density is

$$\rho = \frac{\omega^2 A_0^2}{2\mu_0 c^2} = \tfrac{1}{2}\omega^2\epsilon_0 A_0^2, \tag{9.65}$$

where we have used $\mu_0 c^2 = 1/\epsilon_0$. When we use equation (9.65) to eliminate A_0^2 from equation (9.63), we have our final expression for the rate of radiative decays

$$\begin{aligned} R &= \frac{\pi e^2}{\epsilon_0\hbar^2}|\langle E_k|z|E_N\rangle|^2 \frac{\mathrm{d}\rho}{\mathrm{d}\omega}\bigg|_{\omega=(E_N-E_k)/\hbar} \\ &= \frac{2\pi}{a_0 m_{\rm e}}|\langle E_k|z|E_N\rangle|^2 \frac{\mathrm{d}\rho}{\mathrm{d}\nu}\bigg|_{\nu=(E_N-E_k)/h}. \end{aligned} \tag{9.66}$$

Box 9.1: Einstein A and B Coefficients

In 1916, when only the merest fragments of quantum physics were known, Einstein showed (*Verh. Deutsch. Phys. Ges.* **18**, 318) that systems must be capable of both spontaneous and stimulated emission of photons, and that the coefficient of stimulated emission must equal that for absorption of a photon. He obtained these results by requiring that in thermal equilibrium there are equal rates of absorption and emission of photons of a given frequency ν by an ensemble of systems. He considered a frequency ν for which $h\nu = \Delta E$, the energy difference between two states $|1\rangle$ and $|2\rangle$ of the systems. The rate of absorptions he assumed to be $N_{\rm abs} = B_{\rm a}N_1({\rm d}\rho/{\rm d}\nu)$, where $B_{\rm a}$ is the absorption coefficient, N_1 is the number of systems in the state $|1\rangle$, and $({\rm d}\rho/{\rm d}\nu)$ is the energy density in radiation of frequency ν. The rate of emissions he assumed to be $N_{\rm em} = B_{\rm e}N_2({\rm d}\rho/{\rm d}\nu) + AN_2$, where $B_{\rm e}$ is the coefficient for induced emission and A is that for spontaneous emission. Equating $N_{\rm abs}$ to $N_{\rm em}$ yields

$$0 = (B_{\rm e}N_2 - B_{\rm a}N_1)\frac{{\rm d}\rho}{{\rm d}\nu} + AN_2.$$

In thermal equilibrium $N_1 = N_2{\rm e}^{h\nu/kT}$ and ${\rm d}\rho/{\rm d}\nu$ is given by the Planck function. Using these relations to eliminate N_1 and ${\rm d}\rho/{\rm d}\nu$ and then cancelling N_2, we find

$$0 = (B_{\rm e} - B_{\rm a}{\rm e}^{h\nu/kT})\frac{8\pi h\nu^3}{c^3({\rm e}^{h\nu/kT} - 1)} - A.$$

In the limit of very large T, ${\rm e}^{h\nu/kT} \to 1$, so the factor multiplying the bracket with the Bs becomes large, and the contents of this bracket tends to $B_{\rm e} - B_{\rm a}$. It follows that these coefficients must be equal. We therefore drop the subscripts on them, take B out of the bracket, cancel the factors with exponentials, and finally deduce that

$$A = 8\pi h(\nu/c)^3 B. \tag{1}$$

If we make E_k the initial state and E_N the final state, then the negative-frequency term in equation (9.61) gives rise to excitations at an identical rate. Thus we have recovered from a dynamical argument Einstein's famous result that stimulated emission of photons occurs, and that the coefficient B that controls the rate of stimulated emission is equal to the absorption coefficient (Box 9.1). Einstein was able to relate the coefficient A of spontaneous emission to B. Our treatment does not include spontaneous emission because we have neglected the quantum uncertainty in the electromagnetic field. This uncertainty endows the field with zero-point energy (§3.1), and spontaneous emission can be thought of as emission stimulated by the zero-point energy of the electromagnetic field. Einstein's argument does not yield a numerical value for either A or B. Our quantum mechanical treatment has yielded a value for B, and with Einstein's relation (eq. 1 in Box 9.1) between B and A we can infer the value of A.

$$A = \frac{16\pi^2 h\nu^3}{c^3 a_0 m_{\rm e}}|\langle E_k|z|E_N\rangle|^2. \tag{9.67}$$

From this we can estimate the typical lifetime for radiative decay from an excited state of an atom.

When the radiation density ρ is very small, the number N_2 of atoms in an excited state obeys $\dot{N}_2 = -AN_2$ (Box 9.1), so N_2 decays exponentially with a characteristic time A^{-1}. Unless some symmetry condition causes the matrix element in equation (9.67) to vanish, we expect the value of the matrix element to be $\sim a_0$. So the characteristic radiative lifetime of a state is

$$\tau = A^{-1} = \frac{m_e c^2}{h\nu}\frac{\lambda}{a_0}\frac{1}{16\pi^2\nu}. \tag{9.68}$$

For an optical transition, $h\nu \sim 2\,\mathrm{eV}$, $\lambda \sim 650\,\mathrm{nm} \sim 1.2\times10^4 a_0$, and $\nu \sim 4.6\times10^{14}\,\mathrm{Hz}$, so $\tau \sim 4\times10^{-8}\,\mathrm{s}$. It follows that $\sim 10^7$ oscillations of the atom occur before the radiation of energy causes the atom to slump into the lower state.

9.3.3 Selection rules

Equation (9.66) states that the rate of radiative transitions is proportional to the mod-square of the electric dipole operator ez. For this reason the approximation we made, that $\mathbf{k}\cdot\mathbf{x} \ll 1$, is called the **electric dipole approximation**.

There are important circumstances in which symmetry causes the matrix element of the dipole operator to vanish between the initial and final states. Transitions between such states are said to be **forbidden** in contrast to **allowed transitions**, for which the matrix element does not vanish. Some approximations were involved in our derivation of equation (9.66), so the transition rate does not necessarily vanish completely when the matrix element is zero. In fact, forbidden transitions often *do* occur, but at rates that are much smaller than the characteristic rate of allowed transitions (eq. 9.68) because the rate of a forbidden transition is proportional to terms that we could neglect in the derivation of equation (9.66). We now investigate relations between the initial and final states that must be satisfied if the states are to be connected by an allowed transition. Such relations are called **selection rules**. The slower rate of forbidden transitions must be determined by either including the next term of the Taylor expansion of $\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}$, or taking into account the perturbation $\mu_B \mathbf{S}\cdot\mathbf{B}$ that arises from the interaction of the intrinsic magnetic moment of an electron with the wave's magnetic field.

We are interested in matrix elements between states that are eigenstates of operators that commute with the Hamiltonian H that the atom would have if it were decoupled from electromagnetic waves. The Hamiltonian should include spin-orbit coupling as well as interaction with whatever steady external electric or magnetic fields are being applied. The operator in the matrix element is the component of the position operator parallel to the electric field of the radiation that is being either absorbed or emitted.

Even in the presence of an external field, the angular-momentum parallel to the field, which we may call J_z, commutes with H, so the kets of interest are labelled with m. Since $[J_z, z] = 0$, the ket $z|E,m\rangle$ is an eigenket of J_z with eigenvalue m. It follows that $\langle E,m|z|E',m'\rangle = 0$ unless $m = m'$. This gives us the first selection rule listed in Table 9.1, namely that when the electric vector of the radiation is parallel to the imposed field, the quantum number m is unchanged by radiation.

If we define $x_\pm = x \pm \mathrm{i}y$, we have

$$[J_z, x_\pm] = \mathrm{i}y \pm \mathrm{i}(-\mathrm{i}x) = \pm x_\pm. \tag{9.69}$$

It follows that $x_\pm|E,m\rangle$ is an eigenket of J_z with eigenvalue $m\pm1$, so

$$\begin{aligned}\langle E,m|x|E',m'\rangle &= \tfrac{1}{2}\langle E,m|(x_+ + x_-)|E',m'\rangle \\ &= 0 \text{ unless } m' = m\pm1.\end{aligned} \tag{9.70}$$

Obviously the same result applies to the matrix element for y. Hence we have the second selection rule listed in Table 9.1: when the electric vector of the radiation is perpendicular to the imposed field, the quantum number m changes by ±1. If the direction of observation is along the imposed field, the electric vector of the radiation must be perpendicular to the field. Hence,

Table 9.1 Selection rules

j	$\lvert j-j'\rvert \le 1$ but $j=0 \not\to j'=0$
m	$\lvert m-m'\rvert \le 1$; $m=m'$ for **E** parallel to an external field; $\lvert m-m'\rvert = 1$ for photon emitted parallel to an external field
l	$\lvert l-l'\rvert = 1$
s	$s=s'$

in this case m must change by ±1. In fact, m increases when a left-hand circularly-polarised photon is emitted in the positive z direction, and conversely for the emission of a right-hand polarised photon. When the direction of observation is perpendicular to the imposed field, the electric vector of the radiation can be either perpendicular to the field, in which case m changes by ±1, or parallel to the field, and then m does not change.

When there is no imposed field, m may be unchanged or change by ±1, and we can observe photons associated with any of these changes in m when observing from any direction.

When there is no imposed field, J^2 commutes with H and the kets of interest are labelled with E, j and m. The selection rule for j can be obtained from the rules for adding angular momentum that were discussed in §7.5: $\langle E, j, m|x_k|E', j', m'\rangle$ vanishes unless it is possible to make a spin-j object by adding a spin-one object to a spin-j' object. For example, the matrix element vanishes if $j = j' = 0$ because spin-one is all you can get by adding a spin-one system to a spin-zero one. Subject to the selection rules on m just discussed, the matrix element does not vanish if $j = 0$ and $j' = 1$, or if $j = 1$ and $j' = 1$, because both a spin-zero system and a spin-one one can be obtained by adding two spin-one subsystems. The matrix element vanishes if $j = 1$ and $j' = 3$ because by adding spin-one and spin-three the smallest spin you can get is spin-two. In summary, the selection rule is $|j - j'| \le 1$ except that $j = 0 \to j' = 0$ is forbidden.

The selection rules for j that we have just given follow from a powerful result of group theory, the **Wigner-Eckart theorem**. Unfortunately, a significant amount of group theory is required to prove this theorem. In Appendix G we give a proof of the selection rule for j that builds on the calculation involved in Problem 7.24.

When spin-orbit coupling is weak, the total orbital angular momentum L^2 and the total spin angular momentum S^2 are constants of motion, so their quantum numbers l and s are likely to appear as labels in the kets. Since $[\mathbf{x}, \mathbf{S}] = 0$, it is clear that the selection rule for s is that it should not change. The selection rule for l was derived in Problem 7.24: $|l - l'| = 1$.

Problems

9.1 A harmonic oscillator with mass m and angular frequency ω is perturbed by $\delta H = \epsilon x^2$. What is the exact change in the ground-state energy? What value does perturbation theory give? Hint: use equation (3.18).

9.2 The harmonic oscillator of Problem 9.1 is perturbed by $\delta H = \epsilon x$. Show that the perturbed Hamiltonian can be written

$$H = \frac{1}{2m}\left(p^2 + m^2\omega^2X^2 - \frac{\epsilon^2}{\omega^2}\right),$$

where $X = x + \epsilon/m\omega^2$ and hence deduce the exact change in the ground-state energy. Interpret these results physically.

What value does first-order perturbation theory give? From perturbation theory determine the coefficient b_1 of the unperturbed first-excited state in the perturbed ground state. Discuss your result in relation to the exact ground state of the perturbed oscillator.

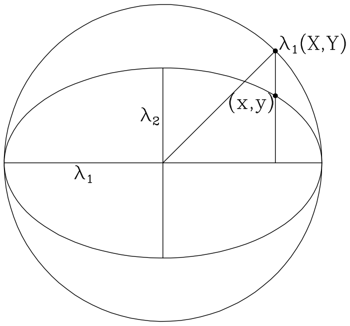


Figure 9.3 The input and output vectors of a 2 × 2 Hermitian matrix are related by a circle with the matrix's largest eigenvalue as radius and the ellipse that has the eigenvalues as sem-axes.

9.3 The harmonic oscillator of Problem 9.1 is perturbed by $\delta H = \epsilon x^4$. Show that the first-order change in the energy of the $n^{\rm th}$ excited state is

$$\delta E = 3(2n^2 + 2n + 1)\epsilon \left(\frac{\hbar}{2m\omega}\right)^2 . \tag{9.71}$$

Hint: use equation (3.18).

9.4 The infinite square-well potential $V(x) = 0$ for $|x| < a$ and ∞ for $|x| > a$ is perturbed by the potential $\delta V = \epsilon x/a$. Show that to first order in ϵ the energy levels of a particle of mass m are unchanged. Show that even to this order the ground-state wavefunction *is* changed to

$$\psi_1(x) = \frac{1}{\sqrt{a}}\cos(\pi x/2a) + \frac{16\epsilon}{\pi^2 E_1\sqrt{a}}\sum_{n=2,4,}(-1)^{n/2}\frac{n}{(n^2-1)^3}\sin(n\pi x/2a),$$

where E_1 is the ground-state energy. Explain physically why this wavefunction does not have well-defined parity but predicts that the particle is more likely to be found on one side of the origin than the other. State with reasons but without further calculation whether the second-order change in the ground-state energy will be positive or negative.

9.5 An atomic nucleus has a finite size, and inside it the electrostatic potential $\Phi(r)$ deviates from $Ze/(4\pi\epsilon r)$. Take the proton's radius to be $a_{\rm p} \simeq 10^{-15}\,$m and its charge density to be uniform. Then treating the difference between Φ and $Ze/(4\pi\epsilon_0 r)$ to be a perturbation on the Hamiltonian (8.11) of hydrogen, calculate the first-order change in the ground-state energy of hydrogen. Why is the change in the energy of any P state extremely small? Comment on how the magnitude of this energy shift varies with Z in hydrogenic ions of charge Z. Hint: exploit the large difference between $a_{\rm p}$ and a_0 to approximate the integral you formally require.

9.6 Evaluate the Landé g factor for the case $l = 1$, $s = \frac{1}{2}$ and relate your result to Figure 9.2.

9.7* The Hamiltonian of a two-state system can be written

$$H = \begin{pmatrix} A_1 + B_1\epsilon & B_2\epsilon \\ B_2\epsilon & A_2 \end{pmatrix}, \tag{9.72}$$

where all quantities are real and ϵ is a small parameter. To first order in ϵ, what are the allowed energies in the cases (a) $A_1 \neq A_2$, and (b) $A_1 = A_2$?

Obtain the exact eigenvalues and recover the results of perturbation theory by expanding in powers of ϵ.

9.8* For the P states of hydrogen, obtain the shift in energy caused by a weak magnetic field (a) by evaluating the Landé g factor, and (b) by use equation (9.27) and the Clebsch–Gordan coefficients calculated in §7.5.2.

9.9 The 2×2 Hermitian matrix $\mathbf{H}$ has positive eigenvalues $\lambda_1 > \lambda_2$. The vectors (X, Y) and (x, y) are related by

$$\mathbf{H} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Show that the points $(\lambda_1 X, \lambda_2 Y)$ and (x, y) are related as shown in Figure 9.3. How does this result generalise to 3×3 matrices? Explain the relation of Rayleigh's theorem to this result.

9.10 Show that for the unnormalised spherically-symmetric wavefunction $\psi(r)$ the expectation value of the gross-structure Hamiltonian of hydrogen is

$$\langle H \rangle = \left(\frac{\hbar^2}{2m_{\rm e}} \int \mathrm{d}r\, r^2 \left| \frac{\mathrm{d}\psi}{\mathrm{d}r} \right|^2 - \frac{e^2}{4\pi\epsilon_0} \int \mathrm{d}r\, r |\psi|^2 \right) \Bigg/ \int \mathrm{d}r\, r^2 |\psi|^2. \qquad (9.73)$$

For the trial wavefunction $\psi_b = \mathrm{e}^{-br}$ show that

$$\langle H \rangle = \frac{\hbar^2 b^2}{2m_{\rm e}} - \frac{e^2 b}{4\pi\epsilon_0},$$

and hence recover the definitions of the Bohr radius and the Rydberg constant.

9.11* Using the result proved in Problem 9.10, show that the trial wavefunction $\psi_b = \mathrm{e}^{-b^2 r^2/2}$ yields $-8/(3\pi)\mathcal{R}$ as an estimate of hydrogen's ground-state energy, where $\mathcal{R}$ is the Rydberg constant.

9.12 Show that the stationary point of $\langle\psi|H|\psi\rangle$ associated with an excited state of H is a saddle point. Hint: consider the state $|\psi\rangle = \cos\theta|k\rangle + \sin\theta|l\rangle$, where θ is a parameter.

9.13* A particle travelling with momentum $p = \hbar k > 0$ from $-\infty$ encounters the steep-sided potential well $V(x) = -V_0 < 0$ for $|x| < a$. Use the Fermi golden rule to show that the probability that a particle will be reflected by the well is

$$P_{\rm reflect} \simeq \frac{V_0^2}{4E^2} \sin^2(2ka),$$

where $E = p^2/2m$. Show that in the limit $E \gg V_0$ this result is consistent with the exact reflection probability derived in §5.3.1.. Hint: adopt periodic boundary conditions so the wavefunctions of the in and out states can be normalised.

9.14* Show that the number states $g(E)\,\mathrm{d}E\,\mathrm{d}^2\Omega$ with energy in $(E, E{+}\mathrm{d}E)$ and momentum in the solid angle $\mathrm{d}^2\Omega$ around $\mathbf{p} = \hbar\mathbf{k}$ of a particle of mass m that moves freely subject to periodic boundary conditions on the walls of a cubical box of side length L is

$$g(E)\,\mathrm{d}E\,\mathrm{d}^2\Omega = \left(\frac{L}{2\pi}\right)^3 \frac{m^{3/2}}{\hbar^3} \sqrt{2E}\,\mathrm{d}E\,\mathrm{d}\Omega^2. \qquad (9.74)$$

Hence show from Fermi's golden rule that the cross section for elastic scattering of such particles by a weak potential $V(\mathbf{x})$ from momentum $\hbar\mathbf{k}$ into the solid angle $\mathrm{d}^2\Omega$ around momentum $\hbar\mathbf{k}'$ is

$$\mathrm{d}\sigma = \frac{m^2}{(2\pi)^2\hbar^4}\,\mathrm{d}^2\Omega \left| \int \mathrm{d}^3\mathbf{x}\, \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} V(\mathbf{x}) \right|^2. \qquad (9.75)$$

Explain in what sense the potential has to be "weak" for this **Born approximation"** to the scattering cross section to be valid.

9.15 From equation (8.59) show that the product a_0k of the Bohr radius and the wavenumber of a photon of energy E satisfies

$$a_0k = \frac{E}{\alpha m_\mathrm{e}c^2}. \tag{9.76}$$

Hence show that the wavenumber k_α of an Hα photon satisfies $a_0k_\alpha = \frac{5}{72}\alpha$ and determine λ_α/a_0. What is the connection between this result and our estimate that $\sim 10^7$ oscillations are required to complete a radiative decay. Does it imply anything about the way the widths of spectral lines from allowed atomic transitions varies with frequency?

9.16 Equation (9.69) implies that $x_\pm$ act as ladder operators for J_z. Why did we not use these operators in §7.1?

9.17 Given that a system's Hamiltonian is of the form

$$H = \frac{p^2}{2m_\mathrm{e}} + V(\mathbf{x}) \tag{9.77}$$

show that $[x,[H,x]] = \hbar^2/m_\mathrm{e}$. By taking the expectation value of this expression in the state $|k\rangle$, show that

$$\sum_{n\neq k} |\langle n|x|k\rangle|^2(E_n - E_k) = \frac{\hbar^2}{2m_\mathrm{e}}, \tag{9.78}$$

where the sum runs over all the other stationary states.

The **oscillator strength** of a radiative transition $|k\rangle \to |n\rangle$ is defined to be

$$f_{kn} \equiv \frac{2m_\mathrm{e}}{\hbar^2}(E_n - E_k)|\langle n|x|k\rangle|^2 \tag{9.79}$$

Show that $\sum_n f_{kn} = 1$. What is the significance of oscillator strengths for the allowed radiative transition rates of atoms?

9.18 At early times ($t \sim -\infty$) a harmonic oscillator of mass m and natural angular frequency ω is in its ground state. A perturbation $\delta H = \mathcal{E}x\mathrm{e}^{-t^2/\tau^2}$ is then applied, where $\mathcal{E}$ and τ are constants.

a. What is the probability that by late times the oscillator transitions to its second excited state, $|2\rangle$?

b. Show that the probability that the oscillator transitions to the first excited state, $|1\rangle$, is

$$P = \frac{\pi\mathcal{E}^2\tau^2}{2m\hbar\omega}\mathrm{e}^{-\omega^2\tau^2/2}, \tag{9.80}$$

c. Plot P as a function of τ and comment on its behaviour as $\omega\tau \to 0$ and $\omega\tau \to \infty$.

10

Helium and the periodic table

In this chapter we build on the foundations laid by our study of hydrogen in Chapter 8 to understand how the atoms of heavier elements work. Most of the essential ideas already emerged in Chapter 8. In fact, only one important point of principle needs to be introduced before we can move down the periodic table understanding why elements have the spectra and chemistry that they do. This outstanding issue is the remarkable implications that quantum mechanics carries for the way in which identical particles interact with one another. We shall be concerned with the case in which the particles are electrons, but the principles we elucidate apply much more widely, for example, to the dynamics of the three quarks that make up a proton or neutron, or the two oxygen atoms that comprise an oxygen molecule.

10.1 Identical particles

Consider a system that contains two identical spinless particles. Then a complete set of amplitudes is given by a function ψ of the coordinates $\mathbf{x}$ and $\mathbf{x}'$ of the particles: the complex number $\psi(\mathbf{x},\mathbf{x}')$ is the amplitude to find one particle at $\mathbf{x}$ and the other particle at $\mathbf{x}'$. What's the physical interpretation of the number $\psi(\mathbf{x}',\mathbf{x})$? It also gives the amplitude to find particles at $\mathbf{x}$ and $\mathbf{x}'$. If the particles were not identical – if one where a pion the other a kaon, for example – finding the pion at $\mathbf{x}$ and the kaon at $\mathbf{x}'$ would be a physically distinct situation from finding the kaon at $\mathbf{x}$ and the pion at $\mathbf{x}'$. But if both particles are pions, $\psi(\mathbf{x},\mathbf{x}')$ and $\psi(\mathbf{x}',\mathbf{x})$ are amplitudes for identical physical situations. Does it follow that $\psi(\mathbf{x},\mathbf{x}') = \psi(\mathbf{x}',\mathbf{x})$? No, because experimentally we can only test the probabilities to which these amplitudes give rise. So we can only safely argue that the probability $|\psi(\mathbf{x},\mathbf{x}')|^2$ must equal the probability $|\psi(\mathbf{x}',\mathbf{x})|^2$, or equivalently that

$$\psi(\mathbf{x},\mathbf{x}') = \mathrm{e}^{\mathrm{i}\phi}\psi(\mathbf{x}',\mathbf{x}), \tag{10.1}$$

where ϕ is a real number. This equation must hold for all $\mathbf{x}$ and $\mathbf{x}'$. So the *function* ψ has the property that if you swap its arguments, you increment its phase by ϕ. Specifically

$$\psi(\mathbf{x}',\mathbf{x}) = \mathrm{e}^{\mathrm{i}\phi}\psi(\mathbf{x},\mathbf{x}'). \tag{10.2}$$

Substituting this equation into the right side of equation (10.1), it follows that

$$\psi(\mathbf{x},\mathbf{x}') = \mathrm{e}^{2\mathrm{i}\phi}\psi(\mathbf{x},\mathbf{x}'), \tag{10.3}$$

which implies that either $\phi = 0$ or $\phi = \pi$. Thus we have shown that the wavefunction for a system of two spinless particles has to be either symmetric or antisymmetric under interchange of the particles' coordinates.

Consider now the case of two spin-s particles, which might, for example, be electrons ($s = \frac{1}{2}$) or photons ($s = 1$). A complete set of amplitudes would be the amplitude for one particle to be at $\mathbf{x}$ in the state that has eigenvalue m of S_z, and the other particle to be at $\mathbf{x}'$ with eigenvalue m'. Let the complex number $\psi_{mm'}(\mathbf{x},\mathbf{x}')$ denote this amplitude – that is, let the first subscript on ψ give the orientation of the spin of the particle that is found at the position given by the first argument of ψ. Then the possibly different amplitude $\psi_{m'm}(\mathbf{x}',\mathbf{x})$ is for the identical physical situation. Hence

$$\psi_{m'm}(\mathbf{x}',\mathbf{x}) = \mathrm{e}^{\mathrm{i}\phi}\psi_{mm'}(\mathbf{x},\mathbf{x}'). \tag{10.4}$$

This equation must hold for all m, m' and $\mathbf{x}$, $\mathbf{x}'$. So swapping the subscripts on ψ at the same time as swapping the arguments, is equivalent to multiplying through by $\mathrm{e}^{\mathrm{i}\phi}$. Swapping a second time leads to

$$\psi_{mm'}(\mathbf{x},\mathbf{x}') = \mathrm{e}^{2\mathrm{i}\phi}\psi_{mm'}(\mathbf{x},\mathbf{x}'), \tag{10.5}$$

so, as in the case of spin-zero particles, either $\phi = 0$ or $\phi = \pi$. It turns out that there is no change of sign if the particles are bosons ($s = 0, 1, 2, \ldots$), and there is a change of sign if the particles are fermions ($s = \frac{1}{2}, \frac{3}{2}, \ldots$). That is

$$\psi_{mm'}(\mathbf{x},\mathbf{x}') = \begin{cases} -\psi_{m'm}(\mathbf{x}',\mathbf{x}) & \text{for fermions} \\ +\psi_{m'm}(\mathbf{x}',\mathbf{x}) & \text{for bosons.} \end{cases} \tag{10.6}$$

Generalisation to the case of N identical particles If our system contains N identical fermions, the wavefunction will change its sign when we swap the arguments (both spin quantum numbers and spatial coordinates) associated with any two slots in the wavefunction. Similarly, if the system contains N bosons, the wavefunction will be invariant when we swap the arguments associated with any two slots.

10.1.1 Pauli exclusion principle

An immediate consequence of the wavefunction of fermions being antisymmetric under a swap of its arguments, is that there is zero probability of finding two fermions with their spins oriented in the same way at the same location: since $\psi_{mm}(\mathbf{x},\mathbf{x}) = -\psi_{mm}(\mathbf{x},\mathbf{x})$, the amplitude $\psi_{mm}(\mathbf{x},\mathbf{x})$ must vanish. Since wavefunctions are continuous functions of position, and their spatial derivatives are constrained in magnitude by the particles' momenta, $\psi_{mm}(\mathbf{x},\mathbf{x}')$ can vanish at $\mathbf{x} = \mathbf{x}'$ only if it is small whenever the two arguments are nearly equal. Hence, fermions with similarly oriented spins avoid each other; they are **anticorrelated**. This fact has profound implications for atomic and condensed-matter physics.

If the particles' spins have different orientations, there can be a non-zero amplitude of finding them at the same location: from $\psi_{mm'}(\mathbf{x},\mathbf{x}) = -\psi_{m'm}(\mathbf{x},\mathbf{x})$ it does not follow that the amplitude $\psi_{mm'}(\mathbf{x},\mathbf{x})$ vanishes.

It is often convenient to expand the wavefunction $\psi_{mm'}(\mathbf{x},\mathbf{x}')$ as a sum of products of single-particle wavefunctions $u_n^m(\mathbf{x})$ (§6.1)

$$\psi_{mm'}(\mathbf{x},\mathbf{x}') = \sum_{nn'} a_{mm'}^{nn'} u_n^m(\mathbf{x}) u_{n'}^{m'}(\mathbf{x}'), \tag{10.7}$$

where the complex numbers $a^{nn'}_{mm'}$ are suitable expansion coefficients. When the particles are fermions, inserting this expansion into equation (10.6) and gathering all terms onto one side yields

$$\begin{aligned} 0 &= \sum_{nn'} \left(a^{nn'}_{mm'} u^m_n(\mathbf{x}) u^{m'}_{n'}(\mathbf{x}') + a^{nn'}_{m'm} u^{m'}_n(\mathbf{x}') u^m_{n'}(\mathbf{x}) \right) \\ &= \sum_{nn'} u^m_n(\mathbf{x}) u^{m'}_{n'}(\mathbf{x}') \left(a^{nn'}_{mm'} + a^{n'n}_{m'm} \right), \end{aligned} \tag{10.8}$$

where the second equality holds because we can swap the names of n and n' in the second sum on the first line. Since this equation must hold for all values of $\mathbf{x}$ and $\mathbf{x}'$, the coefficient of each function in the expansion must vanish separately, and we have

$$0 = a^{nn'}_{mm'} + a^{n'n}_{m'm}. \tag{10.9}$$

Putting $n' = n$ and $m = m'$ we find that $a^{nn}_{mm} = 0$. Thus there is zero amplitude to find that both fermions are in the same single-particle state $u^m_n(\mathbf{x})$. This result is known as the **Pauli exclusion principle**.

The Pauli exclusion principle ensures that any expansion of the form (10.7) involves at least two terms: either $n' \neq n$ or $m' \neq m$, or both. When there are only two terms, equation (10.9) ensures that $a^{nn'}_{mm'} = -a^{n'n}_{m'm}$, so equation (10.7) reduces to

$$\psi_{mm'}(\mathbf{x}, \mathbf{x}') = \text{constant} \times \left\{ u^m_n(\mathbf{x}) u^{m'}_{n'}(\mathbf{x}') - u^m_n(\mathbf{x}') u^{m'}_{n'}(\mathbf{x}) \right\}. \tag{10.10}$$

In §6.1 we saw when the wavefunction is a non-trivial sum over products of a wavefunction for each particle, the particles are correlated. Hence the Pauli exclusion principle implies that identical Fermions are *always* correlated.

It is generally more convenient to work with eigenkets of the total spin operators of the system, rather than the eigenstates $\psi_{mm'}$ of the spin operators of individual particles. In §7.5 we derived the relation between these two sets of basis kets. By far the most important case is that of spin-half particles, when the possible values of m are $\pm\frac{1}{2}$, which we abbreviate to ±. In §7.5.1 we obtained expressions for the spin states of a hydrogen atom when it has (i) spin zero (because the spin of the proton and electron are anti-aligned) and (ii) spin one. We can adapt those expressions to the present case of two electrons by replacing e by $\mathbf{x}$ and p by $\mathbf{x}'$. Then we obtain these expressions for the states ψ^m_s of well defined total spin angular momentum

$$\begin{aligned} \psi^1_1(\mathbf{x}, \mathbf{x}') &= \psi_{++}(\mathbf{x}, \mathbf{x}') \\ \psi^0_1(\mathbf{x}, \mathbf{x}') &= \frac{1}{\sqrt{2}} \{ \psi_{-+}(\mathbf{x}, \mathbf{x}') + \psi_{+-}(\mathbf{x}, \mathbf{x}') \} \\ \psi^{-1}_1(\mathbf{x}, \mathbf{x}') &= \psi_{--}(\mathbf{x}, \mathbf{x}') \end{aligned} \tag{10.11a}$$

and

$$\psi^0_0(\mathbf{x}, \mathbf{x}') = \frac{1}{\sqrt{2}} \{ \psi_{-+}(\mathbf{x}, \mathbf{x}') - \psi_{+-}(\mathbf{x}, \mathbf{x}') \}. \tag{10.11b}$$

The spin-one states are generally called **triplet states** while the spin-zero state is called the **singlet state**. When we exchange $\mathbf{x}$ with $\mathbf{x}'$ and apply the rule (10.6) on the right of each equality, we find that[1]

$$\psi^m_1(\mathbf{x}', \mathbf{x}) = -\psi^m_1(\mathbf{x}, \mathbf{x}') \quad \text{but} \quad \psi^0_0(\mathbf{x}', \mathbf{x}) = +\psi^0_0(\mathbf{x}, \mathbf{x}'). \tag{10.12}$$

---

[1] For example, considering the middle equation of the set (10.11a) we have

$$\psi^0_1(\mathbf{x}, \mathbf{x}') = -2^{-1/2} \{ \psi_{+-}(\mathbf{x}', \mathbf{x}) + \psi_{-+}(\mathbf{x}', \mathbf{x}) \} = -\psi^0_1(\mathbf{x}', \mathbf{x}).$$

That is, the wavefunction of a triplet state is an antisymmetric function of $\mathbf{x}$ and $\mathbf{x}'$, while wavefunction of the singlet state is a symmetric function of $\mathbf{x}$ and $\mathbf{x}'$. We saw above that electrons that have equal components of angular momentum parallel to the z axis avoid each other. Equation (10.12) implies that electrons avoid each other when the projections of their spins into the xy plane are parallel, even when the components parallel to the z axis are anti-aligned.

One way of constructing a function of two variables is to take the product $u(\mathbf{x})v(\mathbf{x}')$ of two functions u and v of one variable. Unless $u = v$, this product is neither symmetric nor antisymmetric under interchange of $\mathbf{x}$ and $\mathbf{x}'$, so it cannot be proportional to the wavefunction of either a triplet or a singlet state. To achieve such proportionality, we must extract the symmetric or antisymmetric part of the product. That is, for appropriate u and v we may have

$$\begin{aligned}\psi_1^m(\mathbf{x},\mathbf{x}') &= \tfrac{1}{2}a_1^m\{u(\mathbf{x})v(\mathbf{x}') - u(\mathbf{x}')v(\mathbf{x})\}\\ \psi_0^0(\mathbf{x},\mathbf{x}') &= \tfrac{1}{2}a_0\{u(\mathbf{x})v(\mathbf{x}') + u(\mathbf{x}')v(\mathbf{x})\},\end{aligned} \tag{10.13}$$

where the complex numbers a_1^m carry the spin dependence of ψ_1^m. In the case $u = v$, the triplet wavefunctions are identically zero but the singlet wavefunction can be non-vanishing; that is, *two* distinct single-particle wavefunctions are required for the construction of a triplet state, while just one single-particle wavefunction is all that is required for a singlet state.

Wavefunctions of the form (10.13) are widely used in atomic physics but one should be clear that it is an approximation to assume that a two-electron wavefunction can be written in terms of just two single-particle wavefunctions; *any* wavefunction can be expanded as a sum of products of single-particle wavefunctions, but the sum will generally contain more than two terms.

10.2 Gross structure of helium

About a quarter of the ordinary matter in the Universe is in the form of helium, the second simplest element after hydrogen. The tools that we now have at our disposal enable us to build a fairly detailed model of these important atoms. This model will illustrate principles that apply in all many-electron atoms.

We seek the stationary states of the Hamiltonian that describes the electrostatic interactions between the two electrons and the alpha particle that make up a helium atom. This Hamiltonian is (cf. eq. 8.1)

$$H = \frac{p_{\rm n}^2}{2m_{\rm n}} + \frac{p_1^2}{2m_{\rm e}} + \frac{p_2^2}{2m_{\rm e}} - \frac{e^2}{4\pi\epsilon_0}\left(\frac{2}{|\mathbf{x}_1 - \mathbf{x}_{\rm n}|} + \frac{2}{|\mathbf{x}_2 - \mathbf{x}_{\rm n}|} - \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|}\right), \tag{10.14}$$

where $\mathbf{x}_i$ and $\mathbf{x}_{\rm n}$ are the position operators of the $i^{\rm th}$ electron and the nucleus, respectively, and $\mathbf{p}_i$ and $\mathbf{p}_{\rm n}$ are the corresponding momentum operators. We shall work in the atom's centre-of-mass frame and neglect the small displacement from the origin and kinetic energy that the nucleus has in this frame. With this approximation, H can be written as the sum of two hydrogenic Hamiltonians with $Z = 2$ (cf. eq. 8.10) and the term that describes the mutual electrostatic repulsion of the electrons

$$H = H_{\rm H}(\mathbf{p}_1,\mathbf{x}_1) + H_{\rm H}(\mathbf{p}_2,\mathbf{x}_2) + \frac{e^2}{4\pi\epsilon_0|\mathbf{x}_1 - \mathbf{x}_2|}, \tag{10.15a}$$

where

$$H_{\rm H}(\mathbf{p},\mathbf{x}) \equiv \frac{p^2}{2m_{\rm e}} - \frac{2e^2}{4\pi\epsilon_0|\mathbf{x}|}. \tag{10.15b}$$

We cannot determine the eigenkets of H exactly, so we resort to the approximate methods developed in the previous chapter.

Box 10.1: Evaluating the integral D_0 in equation (10.20)

We express the two position vectors in spherical polar coordinates. Since $\mathbf{x}_1$ is a fixed vector during the integration over $\mathbf{x}_2$, we are at liberty to orient our z axis for the $\mathbf{x}_2$ coordinate system parallel to $\mathbf{x}_1$. Then $|\mathbf{x}_1 - \mathbf{x}_2| = \sqrt{|r_1^2 + r_2^2 - 2r_1r_2\cos\theta_2|}$ is independent of ϕ_2. The mod square of Ψ_{10}^0 does not depend on ϕ, so the integrand is independent of ϕ_2 and we can trivially integrate over ϕ_2. What remains is

$$D_0 = \frac{2}{a_Z^3}\int \mathrm{d}^3\mathbf{x}_1\,|\Psi_{10}^0(\mathbf{x}_1)|^2 \int \mathrm{d}r_2\mathrm{d}\theta_2 \frac{r_2^2\sin\theta_2 \mathrm{e}^{-2r_2/a_Z}}{\sqrt{|r_1^2 + r_2^2 - 2r_1r_2\cos\theta_2|}} \tag{1}$$

where $a_Z \equiv a_0/2$. Now

$$\frac{\sin\theta_2}{\sqrt{|r_1^2 + r_2^2 - 2r_1r_2\cos\theta_2|}} = \frac{1}{r_1r_2}\frac{\mathrm{d}}{\mathrm{d}\theta_2}\sqrt{|r_1^2 + r_2^2 - 2r_1r_2\cos\theta_2|},$$

so

$$\begin{aligned}\int_0^\pi \frac{\sin\theta_2\,\mathrm{d}\theta_2}{\sqrt{|r_1^2 + r_2^2 - 2r_1r_2\cos\theta_2|}} &= \frac{|r_1+r_2| - |r_1-r_2|}{r_1r_2}\\ &= \begin{cases} 2/r_1 & \text{for } r_1 > r_2\\ 2/r_2 & \text{for } r_1 < r_2.\end{cases}\end{aligned}$$

After using this expression in equation (1), we have to break the integral over r_2 into two parts, and have

$$\begin{aligned}D_0 &= \frac{4}{a_Z^3}\int \mathrm{d}^3\mathbf{x}_1\,|\Psi_{10}^0(\mathbf{x}_1)|^2 \left\{\int_0^{r_1}\mathrm{d}r_2\,\frac{r_2^2}{r_1}\mathrm{e}^{-2r_2/a_Z} + \int_{r_1}^\infty \mathrm{d}r_2\,r_2\mathrm{e}^{-2r_2/a_Z}\right\}\\ &= \frac{2}{a_Z}\int \mathrm{d}^3\mathbf{x}_1\,|\Psi_{10}^0(\mathbf{x}_1)|^2\frac{1}{\rho_1}\left\{2 - \mathrm{e}^{-\rho_1}(2+\rho_1)\right\},\end{aligned}$$

where $\rho_1 \equiv 2r_1/a_Z$. The integral over $\mathbf{x}_1$ is relatively straightforward: given the normalisation of the spherical harmonics, we simply have to integrate over r_1. We transform to the scaled radius ρ_1 and find

$$D_0 = \frac{1}{a_Z}\int \mathrm{d}\rho_1\,\rho_1\mathrm{e}^{-\rho_1}\left\{2 - \mathrm{e}^{-\rho_1}(2+\rho_1)\right\} = \frac{5}{8a_Z} = \frac{5}{4a_0}.$$

10.2.1 Gross structure from perturbation theory

Our first approach is to use the perturbation theory of §9.1. In §8.1 we found the eigenfunctions of H_H. These proved to be products $u_n^l(r)\mathrm{Y}_l^m(\theta,\phi)$ of the radial eigenfunctions u_n^l derived in §8.1.2 and the spherical harmonics Y_l^m derived in §7.2.3. From the work of §6.1 it follows that the eigenfunctions of the operator

$$H_0 \equiv H_\mathrm{H}(\mathbf{p}_1,\mathbf{x}_1) + H_\mathrm{H}(\mathbf{p}_2,\mathbf{x}_2) \tag{10.16}$$

are products

$$\Psi_{nl}^m(\mathbf{x}_1)\Psi_{n'l'}^{m'}(\mathbf{x}_2) \equiv u_n^l(r_1)\mathrm{Y}_l^m(\theta_1,\phi_1)u_{n'}^{l'}(r_2)\mathrm{Y}_{l'}^{m'}(\theta_2,\phi_2), \tag{10.17}$$

where n and n' are any positive integers. From equation (8.22) the corresponding eigenvalues are

$$E_0 \equiv -4\mathcal{R}\left(\frac{1}{n^2} + \frac{1}{n'^2}\right). \tag{10.18}$$

The ground-state wavefunction of H_0 will be a product of the ground-state eigenfunctions $(4\pi)^{-1/2}u_1^0(r)$ of H_H, where the function u_1^0 is given by equation (8.29) with $Z = 2$. From equation (10.18) the ground-state energy of H_0 is

$$E_0 = -8\mathcal{R} = -108.8\,\mathrm{eV}. \tag{10.19}$$

The Hamiltonian (10.14) commutes with all spin operators because it makes no reference to spin. Therefore we are at liberty to seek eigenfunctions of H that are simultaneously eigenfunctions of the total spin operators S^2 and S_z. We have seen that these eigenfunctions are either singlet or triplet states and are either symmetric or antisymmetric functions of the spatial variables. The ground-state wavefunction of H_0 is an inherently a symmetric function of $\mathbf{x}_1$ and $\mathbf{x}_2$, so the ground-state is a singlet. The first-order contribution to the ground-state energy is the expectation value of the perturbing part of the Hamiltonian (10.15.) This expectation value is

$$\Delta E = \frac{e^2}{4\pi\epsilon_0} D_0 \quad \text{where} \quad D_0 \equiv \int \mathrm{d}^3\mathbf{x}_1\, \mathrm{d}^3\mathbf{x}_2\, \frac{|\Psi^0_{10}(\mathbf{x}_1)|^2 |\Psi^0_{10}(\mathbf{x}_2)|^2}{|\mathbf{x}_1 - \mathbf{x}_2|}. \quad (10.20)$$

Box 10.1 describes the evaluation of the six-dimensional integral D_0. We find that $\Delta E = \frac{5}{2}\mathcal{R}$, so our estimate of the ground-state energy of helium is $E = E_0 + \Delta E = -\frac{11}{2}\mathcal{R} = -74.8\,\mathrm{eV}$. The experimentally measured value is $-79.0\,\mathrm{eV}$.

10.2.2 Application of the variational principle to helium

We can use the variational principle (§9.2) refine our estimate of helium's ground-state energy. Our estimate is based on the assumption that the electrons' wavefunctions are those that would be appropriate if the electrons did not repel one another. Suppose we could somehow switch off this repulsion without affecting the attraction between each electron and the alpha particle. Then the electrons would settle into the wavefunctions we have assumed. If we then turned the electric repulsion back on, it would push the electrons apart to some extent, and the atom would become bigger. This thought experiment suggests that we might be able to obtain a better approximation to the electrons' wavefunctions by increasing the characteristic lengthscale that appears in the exponential of a hydrogenic ground-state wavefunction (eq. 8.29) from a_0 to some value a. The variational principle assures us that the minimum value with respect to a of the expectation value of H that we obtain with these wavefunctions will be a better estimate of the ground-state energy than the estimate we obtained by first-order perturbation theory.

Consider, therefore, the expectation value of the Hamiltonian (10.15a) for the case in which the electronic wavefunction is a product of hydrogenic wavefunctions with a_0 replaced by a. From the work of the last subsection we already know the value taken when $a = a_0$. Moreover, the expectation value is made up of five terms, two kinetic energies, and three potential energies, and by dimensional analysis it is clear how each term must scale with a: the kinetic energies scale as a^{-2} because p is proportional to the gradient of the wavefunction, which scales like a^{-1}, while the potential energy contributions scale as a^{-1} since they explicitly have distances in their denominators. We know that when $a = a_0$ and the wavefunctions are hydrogenic, the expectation value of the sum of hydrogenic Hamiltonians in equation (10.15a) is $-8\mathcal{R}$, and we know from the virial theorem (eq. 2.83) that this overall energy is made up of $8\mathcal{R}$ kinetic energy and $-16\mathcal{R}$ of potential energy. We saw above that when $a = a_0$ the electrostatic repulsion of the electrons contributes $\frac{5}{2}\mathcal{R}$ of potential energy. Bearing in mind the way that these kinetic and potential energies scale with a it follows that for general a, the expectation value of helium's Hamiltonian is

$$\langle H \rangle_a = \mathcal{R}\{8x^2 - (16 - \tfrac{5}{2})x\} \quad \text{where} \quad x \equiv \frac{a_0}{a}. \quad (10.21)$$

The derivative of $\langle H \rangle_a$ with respect to x vanishes when $x = \frac{27}{32}$. When we insert this value of x into equation (10.21) we find our improved estimate of helium's ground-state energy is $-\frac{1}{2}(3/2)^6\mathcal{R} = 77.4\,\mathrm{eV}$. As was inevitable, this value is larger than the experimentally measured value of $-79.0\,\mathrm{eV}$. But it is significantly closer than the value we obtained by first-order perturbation theory.

An important indicator of the chemical nature of an element is the magnitude of the energy required to strip an electron from an atom, which is called the element's **ionisation energy**. In the case of hydrogen, this energy is simply the binding energy, 13.6 eV. In the case of helium it is the difference between the binding energies of the atom and the ion He^+ that remains after one electron is stripped away. Since the He^+ ion is hydrogenic with $Z = 2$, its binding energy is $4\mathcal{R} = 54.4\,\mathrm{eV}$, so the ionisation energy of helium is $79.0 - 54.4 = 24.6\,\mathrm{eV}$. This proves to be the largest ionisation energy of any atom, which makes helium perhaps the least chemically active element there is.

10.2.3 Excited states of helium

We now consider the low-lying excited states of helium. Given our success in calculating the ground-state energy of helium with the aid of hydrogenic wavefunctions, it is natural to think about the excited states using the same hydrogenic language. Thus we suppose that the electronic wavefunction is made up of products of single-particle wavefunctions. We recognise that the single-particle wavefunctions that should be used in these products will differ slightly from hydrogenic ones, but we assume that they are similar to the hydrogenic ones that carry the same orbital angular momentum and have the same number, $n-1$, of radial nodes. Hence we can enumerate the single-particle wavefunctions by assigning the usual quantum numbers n and l to each electron. We expect to be able to obtain reasonable estimates of the energies of excited states by taking the expectation value of the Hamiltonian for hydrogenic states.

In the first excited state it is clear that one of the electrons will have been promoted from its $n = 1$ ground state to one of the $n = 2$ states. From our discussion of shielding in §8.1.3, we expect that the state with $l = 0$ will have less energy than any other $n = 2$ state. Thus we seek to construct the first excited state from a product of the hydrogenic ground-state wavefunction $\Psi_1(\mathbf{x})$ and the wavefunction $\Psi_2(\mathbf{x})$ for $n = 2$, $l = 0$, which is also spherically symmetric. Since we are working with distinct single-particle states, we can construct both singlet and triplet states – see equation (10.13). Since the two possibilities differ only in a sign, we defer choosing between them and make our formulae valid for either case, putting the sign for the singlet state on top. We now have to calculate

$$\begin{aligned}\langle H\rangle = \tfrac{1}{2}\int \mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{x}'\ &\{\Psi_1^*(\mathbf{x})\Psi_2^*(\mathbf{x}') \pm \Psi_1^*(\mathbf{x}')\Psi_2^*(\mathbf{x})\}\\ &\times H\,\{\Psi_1(\mathbf{x})\Psi_2(\mathbf{x}') \pm \Psi_1(\mathbf{x}')\Psi_2(\mathbf{x})\}\,.\end{aligned} \tag{10.22}$$

When we substitute for H from equation (10.15a), integrals over terms such as $\Psi_1(\mathbf{x})\Psi_2(\mathbf{x}')H_{\mathrm{H}}\Psi_1(\mathbf{x}')\Psi_2(\mathbf{x})$ arise, where H_{H} is the hydrogenic operator that appears in equations (10.15). The orthogonality of Ψ_1 and Ψ_2 causes these integrals to vanish, because H_{H} contains either $\mathbf{x}$ or $\mathbf{x}'$, but not both operators so there is always an integral of the form $0 = \int \mathrm{d}^3\mathbf{x}\,\Psi_1(\mathbf{x})\Psi_2(\mathbf{x})$. The integral over $\Psi_1(\mathbf{x})\Psi_2(\mathbf{x}')H_{\mathrm{H}}\Psi_1(\mathbf{x})\Psi_2(\mathbf{x}')$ evaluates to either $-4\mathcal{R}$ or $-\mathcal{R}$ depending on whether H_{H} contains $\mathbf{x}$ or $\mathbf{x}'$. Hence

$$\langle H\rangle = -5\mathcal{R} + \frac{e^2}{4\pi\epsilon_0}\{D \pm E\}, \tag{10.23a}$$

where D and E are, respectively, the **direct** and **exchange** integrals:

$$\begin{aligned}D &\equiv \int \mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{x}'\,\frac{|\Psi_1(\mathbf{x})\Psi_2(\mathbf{x}')|^2}{|\mathbf{x}-\mathbf{x}'|}\\ E &\equiv \int \mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{x}'\,\frac{\Psi_1(\mathbf{x})^*\Psi_2(\mathbf{x})\Psi_1(\mathbf{x}')\Psi_2^*(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}.\end{aligned} \tag{10.23b}$$

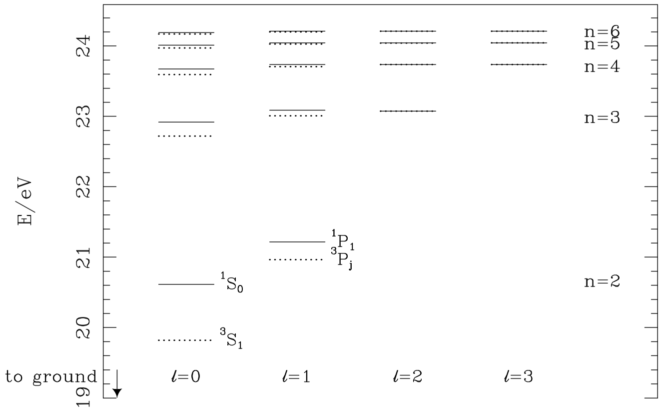


Figure 10.1 Excited states of helium for $n \leq 6$ and $l \leq 3$. Energies are given with respect to the ground-state energy, and the line at the top shows the ionisation energy. Full lines show singlet states and dotted lines show triplet states. Fine structure splits the triplet states with $l \geq 1$ but the splittings are much too small to show on this scale – the largest is 0.00012 eV.

Since both Ψ_1 and Ψ_2 are spherically symmetric functions of their arguments, both integrals can be evaluated by the technique described in Box 10.1. After a good deal of tedious algebra one discovers that $\langle H \rangle = -(56.6 \mp 1.2)\,\mathrm{eV}$, where the upper sign is for the singlet state. The experimentally measured values are $-(58.8 \mp 0.4)\,\mathrm{eV}$. Hence perturbation theory correctly predicts that the triplet state lies below the singlet state.

The differences between our perturbative values and the experimental results arise because the hydrogenic wavefunctions we have employed are not well suited to helium. The deficiency is particularly marked in the case of the $n = 2$ wavefunction because the nuclear charge is significantly shielded from the outer electron, so the $n = 2$ wavefunction should extend to larger radii than the hydrogenic wavefunction we have employed, which assumes that the electron sees the full nuclear charge. Consequently, we have overestimated the **overlap** between the two wavefunctions: the extent to which the wavefunctions permit the electrons to visit the same place. Because our wavefunctions have unrealistically large overlap, they yield values for both D and E that are too large. The exchange integral is particularly sensitive to overestimation of the overlap because it vanishes when there is no overlap, which D does not. Thus it is entirely understandable that our treatment yields binding energies that are insufficiently negative, and a singlet-triplet splitting that is too large.

The sensitivity of the singlet-triplet splitting to wavefunction overlap leaves a clear imprint on the energy-level structure of helium that is shown in Figure 10.1: the separations of corresponding full (singlet) and dotted (triplet) lines diminishes as one goes up any column (increasing n) or from left to right (increasing l). Quantitatively, the singlet-triplet splitting when the excited electron is in a $n = 2$, $l = 1$ state (bottom of second column), rather than the $n = 2$, $l = 0$ state that we have just investigated (bottom of the first column), is 0.2 eV rather than 0.8 eV because, as we saw in §8.1.2, the $l = 1$ state has smaller amplitudes at the small radii at which the $n = 1$ state has large amplitudes.

We have discussed the splitting between singlet and triplet states in the case in which both the single-particle wavefunctions employed are spherically symmetric, so the wavefunctions are entirely real. The analysis for wavefunctions that have $l \neq 0$ is significantly more involved, but the essential result is the same because the exchange integral E is always real (Problem 10.1) and positive. We can see that E is positive as follows. The exchange integral is dominated by the region $\mathbf{x} \simeq \mathbf{x}'$ in which the denominator is small. In this region the numerator does not differ much from $|\Psi_1(\mathbf{x})\Psi_2(\mathbf{x})|^2$, so it is positive. Hence E is positive. Thus it is quite generally true that the triplet states lie below the corresponding singlet state.

In our discussion of spin-orbit coupling in §8.2.1 we saw that the energy scale of that coupling is $\sim \frac{1}{4}Z^2\alpha^4 m_e c^2$ (eq. 8.72b). For helium this evaluates to $\sim 0.006\,\text{eV}$, which is two orders of magnitude smaller than the singlet-triplet splitting. Moreover, we found that the coupling vanishes for states with $l = 0$, so it should vanish in the first excited state of helium. The singlet-triplet splitting is large because it has an electrostatic origin, rather than being a mere relativistic effect: a triplet state has less energy because in it the electrons are anticorrelated (§10.1.1).

It is commonly stated that on account of this anticorrelation the energy of electrostatic repulsion between the electrons is smaller in triplet than in singlet states. This is false: the inter-electron potential energy is *larger* in the triplet than the singlet state.[2] The reason the triplet has lower energy is because it places the electrons closer to the nucleus than the singlet does. Moving the electrons towards the nucleus and thus towards one another naturally increases the energy of electron-electron repulsion, but this increase is outweighed by the lowering of the negative electron-nucleus energy. The quantitative results we obtained above should not be used to evaluate the inter-electron energy because they are based on hydrogenic wavefunctions, which provide a poor approximation to the true wavefunction. As we saw in §9.2, even a poor approximation to the wavefunction of a stationary state yields a useful approximation to the energy of that state because the expectation value of H is stationary in the neighbourhood of a stationary state. But the expectation value of a single term in H, such as the inter-electron potential energy, is *not* extremised by a stationary state, so the error in it will be of order the error in the wavefunction. In particular, to obtain a value that is accurate to first order in the perturbation, it is mandatory to use a wavefunction that is correct to first order, whereas we used the zeroth-order wavefunctions. Because the electrons do a better job of keeping out of each other's way in the triplet state, in that state they can cohabit in a smaller volume, where the attraction of the nucleus is stronger. On account of this effect, the true singlet and triplet wavefunctions differ by more than just a change of sign; in equation (10.22) the functions Ψ_1 and Ψ_2 should also change between the singlet and triplet cases.

The singlet–triplet splitting in helium reflects destructive interference between the amplitudes for the two electrons to be simultaneously at the same place, and it is very much a quantum-mechanical effect. Throughout the periodic table, this mechanism gives rise to large energy differences between atomic states that differ only in their spin. These differences make ferromagnetism possible, and thus provide us with the dynamos, power transformers and electric motors that keep our civilisation on the move.

10.2.4 Electronic configurations and spectroscopic terms

The ground state of helium has neither spin nor orbital angular momentum. In conventional spectroscopic notation (Box 10.2) it is designated $1s^2$, which implies that it has two electrons in the $n = 1$ S state. A related notation is used to indicate the spin, orbital and total angular momentum of the entire atom. In this system the ground state is designated ${}^1S_0$. The superscript 1 implies that the state is a spin-singlet because there is zero spin angular momentum. The S implies that there is no orbital angular momentum, and the subscript 0 implies that there is zero total angular momentum.

The lowest dotted line in Figure 10.1 represents a triplet of excited states. These have the **electronic configuration** $1s2s$ because there is an electron in the $n = 1$, $l = 0$ state and one in the $n = 2$, $s = 0$ state. They

---

[2] B. Schiff, H. Lifson, C.L. Pekeris & P. Rabinowitz, *Phys. Rev.*, **140**, A1104, (1965) find the inter-electron energy to be 6.80 eV in the singlet state and 7.29 eV in the triplet state.

Box 10.2: Spectroscopic Notation

Standard spectroscopic notation presumes that l and s, the total orbital and spin angular momenta, are good quantum numbers. The **electronic configuration** is a specification of the principal (n) and orbital angular momentum (l) quantum numbers of the individual electrons of the outermost shell. Within a configuration a **spectroscopic term** specifies definite values for the total orbital l and spin s angular momenta of the outer electrons. Within each term a **fine-structure level** specifies a definite value for the total electronic angular momentum j. Within a fine-structure level may be distinguished different **hyperfine levels** that differ in total angular momentum f. The letters S, P, D, F denote $l = 0, 1, 2, 3$, respectively.

A typical configuration is denoted $2s2p^3$ meaning one electron has $n = 2$, $l = 0$, and three electrons have $n = 2$, $l = 1$.

Terms are denoted by ${}^{(2s+1)}l_j$; for example ${}^4P_{1/2}$ means $s = \frac{3}{2}$, $l = 1$, $j = \frac{1}{2}$.

form the **spectroscopic term** ${}^3S_1$ because and the angular momenta of the whole atom are given by $s = 1$, $l = 0$ and $j = 1$.

Just above this triplet of states comes the singlet state that has the same electronic configuration $1s2s$ but which forms the distinct spectroscopic term ${}^1S_0$.

Next come four spectroscopic terms that both have the electronic configuration $1s2p$: the most energetic of these terms is the singlet ${}^1P_1$, which is a set of three quantum states that have exactly the same energy but different orientations of the one unit of total angular momentum. Below this are three terms that have very similar energies: ${}^3P_0$, ${}^3P_1$ and ${}^3P_2$. These terms differ from one another in the degrees of alignment of the spin and orbital angular momenta. In the ${}^3P_0$ term the angular momenta are anti-parallel, with the result that the atom has zero angular momentum overall, while in the ${}^3P_2$ term the angular momenta are parallel, so the atom has two units of angular momentum. There is just one quantum state in the ${}^3P_0$ term, and five quantum states in the ${}^3P_2$ term. The small energy differences between the ${}^3P_j$ terms are due to spin-orbit coupling.

Spectrum of helium The selection rules listed in Table 9.1 include $\Delta s = 0$, so in Figure 10.1 transitions between full and dotted levels are forbidden. Hence, an atom which is excited into one of the upper triplet states will cascade down through triplet states until it reaches the ${}^3S_1$ level at the bottom of the triplet hierarchy. The states in this level are **metastable** because they can decay radiatively only by making the forbidden transition to the ${}^1S_0$ ground state, which takes appreciable time. Table 9.1 includes the rule $\Delta l = \pm 1$, so transitions are only allowed between states that lie in adjacent columns, and the excited singlet state that is designated ${}^1S_0$ is also metastable.

10.3 The periodic table

The understanding of atomic structure that we have gained in our studies of hydrogen and helium suffices to explain the structure of the periodic table of the elements.

10.3.1 From lithium to argon

Imagine that we have a helium atom in its first excited state and that we simultaneously add a proton to the nucleus and an electron to the vacancy with principal quantum number $n = 1$ that arose when the atom was put into its excited state. After making these changes we would have a lithium atom in its ground state. The effects on the outermost electron of adding the positively charged proton and the negatively charged electron might be expected to largely cancel, so we would expect the ionisation energy of lithium to be similar to that of a helium atom that's in its first excited state. This expectation is borne out by experimental measurements: the ionisation energy of once excited helium is 4.77 eV while that of lithium in its ground state is 5.39 eV. Thus the energy required to strip an electron from lithium is smaller than that required to take an electron from hydrogen or helium by factors of 2.5 and 4.6, respectively. The comparative ease with which an electron can be removed from a lithium atom makes compounds such as LiH stable (Problem 10.3). It also makes lithium is a metal by making it energetically advantageous for each atom in a lithium crystal to contribute one electron to a common pool of delocalised electrons.

In their ground states atoms of hydrogen and helium cannot absorb radiation at optical frequencies because the first excited states of these atoms lie rather far above the ground state (10.2 and 19.8 eV, respectively). The first excited state of lithium is obtained by promoting the $n = 2$ electron from $l = 0$ to $l = 1$. This change in quantum numbers only increases the electron's energy by virtue of shielding (§8.1.3), so the energy difference is a mere 1.85 eV, the quantity of energy carried by photons of wavelength 671 nm that lie towards the red end of the optical spectrum. Elements that lie beyond helium in the periodic table, the so-called **heavy elements**, feature very prominently in astronomical measurements even though they are present in trace amounts compared to hydrogen and helium because their absorption spectra contain lines at easily observed optical wavelengths.

There is a useful parallel between a lithium atom and a hydrogen atom in its first excited state: the lithium nucleus, shielded by the two $n = 1$ electrons, appears to have the same net charge as the proton in hydrogen, so the $n = 2$ electron moves in a similar electric field to that experienced by an electron with $n = 2$ in hydrogen. We can test this parallel quantitatively by comparing the ionisation energy of lithium (5.39 eV) with the energy of H with $n = 2$ (3.40 eV). This agreement is not terribly good because the $n = 2, l = 0$ wavefunction that forms the ground state of lithium overlaps significantly with the $n = 1$ wavefunction, and therefore has exposure to the full nuclear charge. There is a more satisfying parallel between the first excited state of lithium, in which the $n = 2$ electron has $l = 1$ and the corresponding state of hydrogen: in this state lithium has ionisation energy 3.54 eV.

Consider now the effect of transmuting lithium into beryllium by simultaneously increasing the nuclear charge by one unit and adding a second electron to the $n = 2$, $l = 0$ state. The parallel that we have just described suggests that this operation will be analogous to moving up from hydrogen to helium, and will significantly increase the ionisation energy of the atom. Experiment bears out this expectation, for the ionisation energy of beryllium is 9.32 eV, 1.7 times that of lithium. As in helium, the ground state of beryllium has total spin zero, while the first excited states have spin one. However, whereas in the excited states of helium the two electrons have different values of n, in beryllium they both have $n = 2$, and they differ in their values of l. Consequently, the overlap between the single-electron states that form the beryllium triplet is significantly larger than the corresponding overlap in helium. This fact makes the exchange integral in equations (10.23) large and causes the singlet excited state to lie 2.5 eV above the triplet of excited states.

If we add a unit of charge to the nucleus of a beryllium atom, we create an atom singly ionised boron. The four electrons with $l = 0$ that envelop the ion's nucleus screen the nuclear charge to a considerable extent from the perspective of the lowest-energy unfilled single-particle state, which is a $2p$

H^{1} $1s^1$ $^2S_{1/2}$																	He^{2} $1s^2$ $^1S_0$
Li^{3} $2s^1$ $^2S_{1/2}$	Be^{4} $2s^2$ $^1S_0$											B^{5} $2p^1$ $^2P_{1/2}$	C^{6} $2p^2$ $^3P_0$	N^{7} $2p^3$ $^4S_{3/2}$	O^{8} $2p^4$ $^3P_2$	F^{9} $2p^5$ $^2P_{3/2}$	Ne^{10} $2p^6$ $^1S_0$
Na^{11} $3s^1$ $^2S_{1/2}$	Mg^{12} $3s^2$ $^1S_0$											Al^{13} $3p^1$ $^2P_{1/2}$	Si^{14} $3p^2$ $^3P_0$	P^{15} $3p^3$ $^4S_{3/2}$	S^{16} $3p^4$ $^3P_2$	Cl^{17} $3p^5$ $^2P_{3/2}$	Ar^{18} $3p^6$ $^1S_0$
K^{19} $4s^1$ $^2S_{1/2}$	Ca^{20} $4s^2$ $^1S_0$	Sc^{21} $3d^1$ $^2D_{3/2}$	Ti^{22} $3d^2$ $^3F_2$	V^{23} $3d^3$ $^4F_{3/2}$	Cr^{24} $4s^1 3d^5$ $^7S_3$	Mn^{25} $3d^5$ $^6S_{5/2}$	Fe^{26} $3d^6$ $^5D_4$	Co^{27} $3d^7$ $^4F_{9/2}$	Ni^{28} $3d^8$ $^3F_4$	Cu^{29} $4s^1 3d^{10}$ $^2S_{1/2}$	Zn^{30} $3d^{10}$ $^1S_0$	Ga^{31} $4p^1$ $^2P_{1/2}$	Ge^{32} $4p^2$ $^3P_0$	As^{33} $4p^3$ $^4S_{3/2}$	Se^{34} $4p^4$ $^3P_2$	Br^{35} $4p^5$ $^2P_{3/2}$	Kr^{36} $4p^6$ $^1S_0$
Rb^{37} $5s^1$ $^2S_{1/2}$	Sr^{38} $5s^2$ $^1S_0$	Y^{39} $4d^1$ $^2D_{3/2}$	Zr^{40} $4d^2$ $^3F_2$	Nb^{41} $5s^1 4d^4$ $^6D_{1/2}$	Mo^{42} $5s^1 4d^5$ $^7S_3$	Tc^{43} $5s^1 4d^6$ $^6D_{9/2}$	Ru^{44} $5s^1 4d^7$ $^5F_5$	Rh^{45} $5s^1 4d^8$ $^4F_{9/2}$	Pd^{46} $5s^0 4d^{10}$ $^1S_0$	Ag^{47} $5s^1 4d^{10}$ $^2S_{1/2}$	Cd^{48} $4d^{10}$ $^1S_0$	In^{49} $5p^1$ $^2P_{1/2}$	Sn^{50} $5p^2$ $^3P_0$	Sb^{51} $5p^3$ $^4S_{3/2}$	Te^{52} $5p^4$ $^3P_2$	I^{53} $5p^5$ $^2P_{3/2}$	Xe^{54} $5p^6$ $^1S_0$

Figure 9.2 The first five rows of the periodic table.

state ($n = 2$, $l = 1$). The screening is far from complete, however, so the nuclear charge Z that the outermost electron perceives is greater than unity and the dynamics of the outermost electron of boron is similar to that of the electron in a hydrogenic atom with $Z > 1$. The ionisation energy from the $n = 2$ level of hydrogen is $\frac{1}{4}Z^2\mathcal{R} = 3.40Z^2\,\mathrm{eV}$, while that of boron is 8.30 eV, so $Z \sim 1.6$.

Spin-orbit coupling causes the ground state of boron to form the $^2P_{1/2}$ term in which the electron's spin and orbital angular momenta are antiparallel. At this early stage in the periodic table, spin-orbit coupling is weak, so the excited states of the $^2P_{3/2}$ term lie only 0.0019 eV above the ground state. C^+ ions have the same electronic configuration as boron atoms, and in interstellar space are more abundant by factors of several thousand. Even at the low temperatures ($\sim 20\,\mathrm{K}$) that are characteristic of dense interstellar clouds, collisions carry enough energy to lift C^+ ions into the low-lying excited states of the $^2P_{3/2}$ term, so such collisions are often inelastic, in contrast to collisions involving the very much more abundant hydrogen and helium atoms and hydrogen molecules. At the low densities prevalent in interstellar space, an excited C^+ ion usually has time to return to the ground state by emitting a photon before it is involved in another collision. So C^+ ions cool the interstellar gas by radiating away its kinetic energy. As a result of this physics, the temperature of interstellar gas depends sensitively on the abundances in it of the commonest heavy elements, carbon, nitrogen and oxygen. The propensity of interstellar gas to collapse gravitationally into stars depends on the temperature of the gas, so the formation of stars depends crucially on the existence of low-lying excited states in boron and the next few elements in the periodic table.

When we add another unit of charge to the nucleus of a boron atom, the binding energy of the outermost electron increases by a factor of order $(6/5)^2 = 1.44$. Adding a further electron, which can go into another $2p$ state alongside the existing outer electron, offsets this increase in binding energy to some extent, so we expect the ionisation energy of carbon to lie somewhere between the 8.30 eV of boron and 1.44 times this value, 12.0 eV. The experimental value is 11.3 eV, which lies at the upper end of our anticipated range, implying that the mutual repulsion of the two $2p$ electrons is not very important energetically. This is to be expected because the ground state of carbon belongs to the triplet term[3] $^3P_0$, so the electrons keep out of each other's way. As in boron, the first excited states lie very close to the ground state – they form a $^3P_1$ term 0.0020 eV above the ground state, and there is a $^3P_2$ term 0.0033 eV above that.

Adding a unit of charge to the nucleus of carbon and then dropping an electron into another $2p$ single-particle state creates a nitrogen atom. The ionisation energy increases (to 14.5 eV) for exactly the same reason that it did when we transmuted boron into carbon. The spin of all three $2p$ electrons are aligned to ensure that the wavefunction is antisymmetric in the electrons' spatial coordinates. Hence the ground state of nitrogen belongs to a quadruplet of states. The total orbital angular momentum proves to be zero (Problem 10.5), so all states in this quadruplet have the same energy, and there are actually four distinct ground states that together comprise the term $^4S_{3/2}$ – it is sometimes rather confusingly said that the ground 'state' of nitrogen is four-fold degenerate. The lowest excited states form the $^2D_{3/2}$ term, and they lie 2.3835 eV and 2.3846 eV above the ground state.

Since there are only three single-particle spatial wavefunctions available with $l = 1$, namely the wavefunctions for $m = \pm 1, 0$, when we add an electron to nitrogen to form oxygen, the overall wavefunction cannot be totally antisymmetric in the spatial coordinates of the electrons. The result is that

---

[3] See Problem 10.4 for an explanation of why the ground state of carbon has $l = 1$ rather than $l = 0$ or $l = 2$, which are the other possible results of combining two electrons, each of which has $l = 1$.

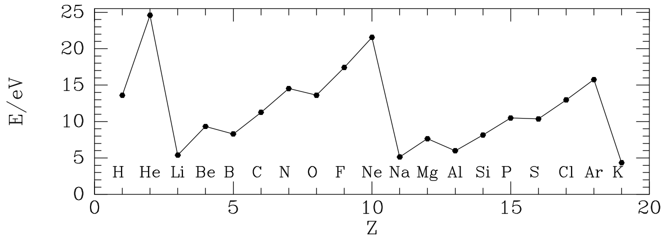


Figure 10.3 Ionisation energies of the first nineteen elements.

in oxygen the electrons are less effective in keeping out of each other's way than are the electrons in carbon and nitrogen, and the ionisation energy of oxygen (13.6 eV) is slightly smaller than that of nitrogen. The ground states form the term $^3P_2$, while states in the $^3P_1$ and $^3P_0$ terms lie 0.020 eV and 0.028 eV above the ground state. In these terms three of the electrons have cancelling orbital angular momenta as in nitrogen, so the orbital angular momentum of the atom is just the single unit introduced by the fourth electron: hence the P in the ground-state term. Spin-orbit interaction causes the ground state to have the largest available value of j, whereas in carbon the reverse was the case.

The easiest way to understand fluorine, the element that follows oxygen in the periodic table, is to skip ahead two places from oxygen to neon, in which a full house of six electrons is packed into the $2p$ states. Every spin is paired and every value of the orbital angular momentum quantum number m is used twice, so both the spin and the orbital angular momenta sum to zero. Each of the six $2p$ electrons is exposed to a large fraction of the ten units of charge on the nucleus, so the ionisation energy is large, 21.6 eV, second only to helium of all the elements. There are no low-lying excited states. This fact together with the large ionisation potential makes neon chemically inactive.

We transmute neon into fluorine by taking away a unit of nuclear charge and one of the $2p$ electrons. The 'hole' we have left in the otherwise complete shell of $2p$ electrons behaves like a spin-half particle that carries one unit of orbital angular momentum. Hence the ground state of fluorine has $s = \frac{1}{2}$ and $l = 1$. Spin-orbit interaction causes the $^2P_{3/2}$ term to lie 0.050 eV below the $^2P_{1/2}$ term that also arises when spin-half is combined with one unit of orbital angular momentum. In the case of oxygen we encountered a similar maximisation of j, and it turns out that the ground states of atoms with shells that are more than half full generally maximise j, while j is minimised in the ground state of an atom with a shell that is less than half full.

We have now reached the end of the first long period of the table. The second long period, from sodium to argon, is an almost perfect copy of the period we have just covered. Figure 10.3 illustrates this statement by showing the ionisation energies of the elements in the first three periods. There is an abrupt drop in the ionisation energy as we move from neon to sodium, from an inert noble gas to a highly reactive alkali metal. Then the ionisation energy creeps up as one moves along the period, with two small setbacks, between magnesium and aluminium, and between phosphorus and sulfur, that are associated with the start and the half-way point of the $3p$, respectively.

10.3.2 The fourth and fifth periods

After reaching argon with its full $3p$ shell, one might expect to start filling the $3d$ shell. Actually the $4s$ states prove to be lower-lying because their vanishing angular momentum allows much greater penetration of the cloud of negative charge associated with the electrons that have $n \leq 3$. But once the $4s$ shell has been filled, filling of the $3d$ shell commences with scandium and continues unbroken through zinc. Once the $3d$ shell is full, filling of the $4p$ shell commences, finishing with the noble gas krypton at $Z = 39$.

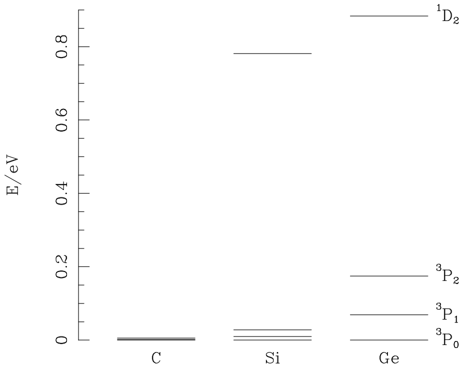


Figure 10.4 The lowest-lying energy levels of carbon silicon and germanium. Along this sequence the fine-structure splitting between the three lowest-lying levels increases dramatically. In the case of germanium the spread in the energies of the triplet states is no longer negligible compared to the energy gap between the lowest-lying singlet state and the top triplet state.

In the next period, filling of the $5s$ shell takes precedence over filling of the $4d$ shell, and when, after cadmium at $Z = 48$, the $4d$ shell is full, filling the $5p$ shell takes precedence over filling the $4f$ shell. The last two periods are very long and have complex patterns due to the availability of shells with large l that tend to be filled much later than shells with the same n but small l.

It is instructive to compare the pattern of energy levels as we move down one column of the periodic table. Figure 10.4 shows the lowest energy levels of carbon and the elements, silicon and germanium, that lie beneath it in the periodic table. As we saw above, carbon has at the bottom of its energy-level diagram a cluster of three very closely spaced energy levels that form the terms ^3P_j for $j = 0, 1, 2$. As we proceed through silicon and germanium the spacing within this cluster grows markedly because it is caused by spin-orbit coupling, which scales like Z^4 (§8.2.1). By the time we reach silicon the energy differences created by spin-orbit coupling are no longer very small compared to the energy difference between the triplet and singlet states, which we know is of electrostatic origin. The total electron spin operator $S^2 \equiv (\sum_i \mathbf{S}_i)^2$ does not commute with the term in the Hamiltonian which is generated by spin-orbit interaction, which is proportional to $\sum_i \mathbf{S}_i \cdot \mathbf{L}_i$. So long as this term is small compared to the terms in the Hamiltonian that do commute with S^2, total electron spin is a good quantum number and it is meaningful to describe the atom with a spectroscopic term such as $^3P_0$. In reality the atom's wavefunction will include contributions from states that have $l \neq 1$, but the coefficients of these terms will be very small, and for most purposes they can be neglected. As the contribution to the Hamiltonian from spin-orbit coupling grows, the coefficients in the ground-state wavefunction of terms with $l \neq 1$ grows, and the designation $^3P_0$ becomes misleading. The lowest-lying three levels of germanium can for most purposes be treated as ^3P_j terms. In the case of tin, which lies under germanium, the designation ^3P_j is highly questionable, and in lead, which lies under tin, it is valueless.

Problems

10.1 Show that the exchange integral defined by equation (10.23b) is real for any single-particle wavefunctions Ψ_1 and Ψ_2.

10.2 The H$^-$ ion consists of two electrons bound to a proton. Estimate its ground-state energy by adapting the calculation of helium's ground-state energy that uses the variational principle. Show that the analogue for H$^-$ of equation (10.21) is

$$\langle H \rangle_a = \mathcal{R}(2x^2 - \tfrac{11}{4}x) \quad \text{where} \quad x \equiv \frac{a_0}{a}. \tag{10.24}$$

Hence find that the binding energy of H$^-$ is $\sim 0.945\mathcal{R}$. Will H$^-$ be a stable ion?

10.3* Assume that a LiH molecule comprises a Li$^+$ ion electrostatically bound to an H$^-$ ion, and that in the molecule's ground state the kinetic energies of the ions can be neglected. Let the centres of the two ions be separated by a distance b and calculate the resulting electrostatic binding energy under the assumption that they attract like point charges. Given that the ionisation energy of Li is $0.40\mathcal{R}$ and using the result of Problem 10.2, show that the molecule has less energy than that of well separated hydrogen and lithium atoms for $b < 4.4a_0$. Does this calculation suggest that LiH is a stable molecule? Is it safe to neglect the kinetic energies of the ions within the molecule?

10.4* Two spin-one gyros are a box. Express that states $|j,m\rangle$ in which the box has definite angular momentum as linear combinations of the states $|1,m\rangle|1,m'\rangle$ in which the individual gyros have definite angular momentum. Hence show that

$$|0,0\rangle = \frac{1}{\sqrt{3}}(|1,-1\rangle|1,1\rangle - |1,0\rangle|1,0\rangle + |1,1\rangle|1,-1\rangle)$$

By considering the symmetries of your expressions, explain why the ground state of carbon has $l = 1$ rather than $l = 2$ or 0. What is the total spin angular momentum of a C atom?

10.5* Suppose we have three spin-one gyros in a box. Express the state $|0,0\rangle$ of the box in which it has no angular momentum as a linear combination of the states $|1,m\rangle|1,m'\rangle|1,m''\rangle$ in which the individual gyros have well-defined angular momenta. Hint: start with just two gyros in the box, giving states $|j,m\rangle$ of the box, and argue that only for a single value of j will it be possible to get $|0,0\rangle$ by adding the third gyro; use results from Problem 10.4.

Explain the relevance of your result to the fact that the ground state of nitrogen has $l = 0$. Deduce the value of the total electron spin of an N atom.

10.6* Consider a system made of three spin-half particles with individual spin states $|\pm\rangle$. Write down a linear combination of states such as $|+\rangle|+\rangle|-\rangle$ (with two spins up and one down) that is symmetric under any exchange of spin eigenvalues $\pm$. Write down three other totally symmetric states and say what total spin your states correspond to.

Show that it is not possible to construct a linear combination of products of $|\pm\rangle$ which is totally antisymmetric.

What consequences do these results have for the structure of atoms such as nitrogen that have three valence electrons.

There are just three of these product states to consider because there

11

Adiabatic principle

We often need to understand the quantum mechanics of systems that have a large number of degrees of freedom. We might. for example, be interested in the speed at which sound waves propagate through a macroscopic crystal of diamond. This depends on the deformability of the bonds between the crystal's carbon atoms, which is in turn determined by the orbits of the atoms' electrons. Also relevant is the inertia of a block of crystal, which is mostly contributed by the carbon nuclei. These nuclei are dynamical systems, in which protons and neutrons move at mildly relativistic speed. Each proton or neutron is itself a dynamical systems in which three quarks and some gluons race about relativistically. When a sound wave passes through the crystal, each nucleus experiences accelerations that must affect its internal dynamics, and the dynamics of its constituent quarks. Is there any chance that a sound wave will induce a nucleus to transition to an excited state? Could a sound wave cause an atom to become electronically excited?

So long as such transitions are realistic possibilities, it is going to be extremely difficult to calculate the speed of sound, because the calculation is going to involve atomic physics, nuclear physics and quantum chromodynamics (**QCD**). The adiabatic approximation, which is the subject of this chapter, enables us to infer that such transitions are exceedingly unlikely to occur. Consequently, in this case and a vast number of similar situations, the adiabatic approximation greatly simplifies our problem by permitting us to neglect phenomena, such as electron or nuclear excitation, that have energy scales that are significantly larger than the characteristic energy scale of the phenomenon under investigation, even though the different degrees of freedom are dynamically coupled. Moreover, we shall see that the adiabatic approximation enables us to calculate quantities such as the spring constant of the bonds that bind a crystals's atoms from the dynamics of the electrons that form these bonds. It also provides the theoretical underpinning for the kinetic theory of gases, for most of condensed-matter physics and much of chemistry. It is enormously important for the development of quantum field theory and our understanding of QCD. Hence, the adiabatic approximation is an extraordinarily important tool with applications that span the natural sciences.

We start by deriving the adiabatic approximation. Then we study in turn elementary applications of it to kinetic theory, to thermodynamics, to condensed-matter physics, and to chemistry.

11.1 Derivation of the adiabatic principle

In §2.2 we stressed that the TDSE (2.26) is valid even when H is time-dependent. However, we have mostly confined ourselves to Hamiltonians that are time-independent. In §9.3 we did consider a time-dependent Hamiltonian, but we assumed that the time-dependent part of H was small. Now we consider the case in which H can change by an amount that is large, so long as the time T over which this change takes place is long in a sense that will be specified below.

We consider the dynamics of a system that has a slowly varying Hamiltonian $H(t)$. At any instant, $H(t)$ has a complete set of eigenkets $|E_n(t)\rangle$ and eigenvalues $E_n(t)$. For the case of vanishing time dependence, equation (9.42) provides an appropriate trial solution of the TDSE (2.26). After modifying this solution to allow for time-variation of the E_n, we have

$$|\psi,t\rangle = \sum_n a_n(t)\exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\,E_n(t')\right)|E_n(t)\rangle. \tag{11.1}$$

Since for each t the set $\{|E_n(t)\rangle\}$ is complete and the numbers $a_n(t)$ are arbitrary, no assumption is involved in writing down this expansion of the system's ket. When we substitute the expansion (11.1) into the TDSE, we find

$$\begin{aligned}\mathrm{i}\hbar\frac{\partial|\psi\rangle}{\partial t} = H|\psi\rangle &= \sum_n a_n \exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\,E_n(t')\right)H(t)|E_n(t)\rangle \\ &= \sum_a \left(\{\mathrm{i}\hbar\dot{a}_n + a_n E_n(t)\}|E_n(t)\rangle + \mathrm{i}\hbar a_n\frac{\partial|E_n\rangle}{\partial t}\right) \\ &\qquad\qquad \times\exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\,E_n(t')\right).\end{aligned} \tag{11.2}$$

Exploiting the fact that $|E_n(t)\rangle$ is an eigenket of $H(t)$ we can cancel a term from each side and are left with

$$0 = \sum_n \left(\dot{a}_n|E_n(t)\rangle + a_n\frac{\partial|E_n\rangle}{\partial t}\right)\exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\,E_n(t')\right). \tag{11.3}$$

Now we use the perturbation theory developed in §9.1 to expand $|E_n(t+\delta t)\rangle$ as a linear combination of the complete set $\{|E_n(t)\rangle\}$. That is, we write

$$|E_n(t+\delta t)\rangle - |E_n(t)\rangle = \sum_{m\neq n} b_{nm}|E_m(t)\rangle, \tag{11.4}$$

where from (9.9) we have

$$b_{nm} = \frac{\langle E_m(t)|\delta H|E_n(t)\rangle}{E_n(t)-E_m(t)} \tag{11.5}$$

with δH the change in H between t and $t+\delta t$. Dividing equation (11.4) by δt and substituting the result into (11.3) we find

$$0 = \sum_n \left(\dot{a}_n|E_n(t)\rangle + a_n\sum_{m\neq n}\frac{\langle E_m(t)|\dot{H}|E_n(t)\rangle}{E_n(t)-E_m(t)}|E_m\rangle\right)\exp\left(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t'\,E_n(t')\right). \tag{11.6}$$

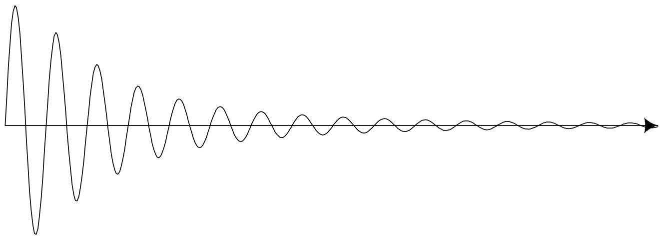


Figure 11.1 A plot of $\sin(kx)$ times the slowly varying function $1/(1+x^2)$. As $k \to \infty$ and the wavelength of the oscillations becomes shorter, the negative contribution from each section of the curve below the x axis more nearly cancels the positive contribution from the preceding upward section.

When we multiply through by $\langle E_k(t)|$ this yields

$$\dot{a}_k = -\sum_{n\neq k} a_n(t) \frac{\langle E_k(t)|\dot{H}|E_n(t)\rangle}{E_n(t) - E_k(t)} \exp\Big(-\frac{\mathrm{i}}{\hbar}\int_0^t \mathrm{d}t' \,\{E_n(t') - E_k(t')\}\Big). \tag{11.7}$$

Although we have used first-order perturbation theory, our working so far has been exact because we can make δH as small as we please by taking δt to be small. Now we introduce an approximation by supposing that H is a slowly varying function of time in the sense that it changes by very little in the time $\hbar/\min(|E_n - E_k|)$, which is the time required for significant motion to occur as a result of interference between the stationary states with energies E_n and E_k (§3.2). In this approximation, the right side of equation (11.7) is a product of a slowly varying function of time and an approximately sinusoidal term that oscillates much more rapidly. When we integrate this expression to get the change in a_k, the integral vanishes rather precisely because the contributions from adjacent half-periods of the oscillating factor nearly cancel (Figure 11.1). Hence, if initially $a_k = 1$ for some k, it will remain unity throughout the evolution. This completes the derivation of the **adiabatic approximation**: *if a system is initially in the $k^{\rm th}$ state of well-defined energy, it will stay in this state when the Hamiltonian is changed sufficiently slowly.*

11.2 Application to kinetic theory

Consider air that is being compressed in the cylinder of a bicycle pump. The air resists the compression by exerting pressure on the cylinder and its piston, and it grows hot as we drive the piston in. This phenomenon is usually explained by treating the air molecules as classical particles that bounce elastically off the cylinder walls. In this section we use the adiabatic principle to interpret the phenomenon at a quantum-mechanical level.

We proceed by first imagining that there is only one molecule in the cylinder, and then making the assumption that when there are a large number N of molecules present, the pressure is simply N times the pressure we calculate for the single-particle case. The Hamiltonian that governs our basic system, a particle in a box, is

$$H(t) = \frac{p^2}{2m} + V(\mathbf{x}, t), \tag{11.8}$$

where the potential $V(\mathbf{x}, t)$ is provided by the walls of the box. The simplest model is

$$V(\mathbf{x}, t) = \begin{cases} 0 & \text{for } \mathbf{x} \text{ in the cylinder} \\ \infty & \text{for } \mathbf{x} \text{ in a wall or the piston.} \end{cases} \tag{11.9}$$

The time dependence of V arises because the piston is moving. We need to find the eigenvalues E_n and eigenkets $|E_n\rangle$ of the Hamiltonian (11.8). We work in the position representation, in which the eigenvalue equation becomes

$$-\frac{\hbar^2}{2m}\nabla^2 u_n + V u_n = E_n u_n \tag{11.10}$$

with $u_n(\mathbf{x}) \equiv \langle \mathbf{x}|E_n\rangle$. From §5.1.1(a) we have that u_n should vanish on the walls of the cylinder and the piston. For $\mathbf{x}$ inside the cylinder, the second term on the left of equation (11.10) vanishes, so E_n and $u_n(\mathbf{x})$ are the solutions to

$$-\frac{\hbar^2}{2m}\nabla^2 u_n = E_n u_n \quad \text{with} \quad u_n = 0 \text{ on the boundary.} \tag{11.11}$$

We assume that the cylinder's cross section is rectangular or circular, so coordinates exist such that (i) the cylinder's walls are all surfaces on which one coordinate vanishes and (ii) the Laplacian operator separates. That is, we can write

$$\nabla^2 = \nabla_2^2 + \frac{\partial^2}{\partial z^2}, \tag{11.12}$$

where ∇_2^2 is an operator that depends on the two coordinates, x and y, that specify location perpendicular to the cylinder's axis, and z is distance down that axis. In this case, we can find a complete set of solutions to equation (11.11) for eigenfunctions that are products $u_n(\mathbf{x}) = X(x,y)Z(z)$ of a function X of x and y, and a function of z alone. Substituting this expression for u_n into equation (11.11) and rearranging, we find

$$Z\nabla_2^2 X + \frac{2mE_n}{\hbar^2}XZ = -X\frac{\mathrm{d}^2 Z}{\mathrm{d}z^2}. \tag{11.13}$$

When we divide through by XZ, we find that the left side does not depend on z while the right side does not depend on x or y. It follows that neither side depends on any of the coordinates. That is, both sides are equal to some constant, which we may call $2m\mathcal{E}_z/\hbar^2$. This observation enables us to separate our original wave equation into two equations

$$\begin{aligned} -\nabla_2^2 X &= \frac{2m(E_n - \mathcal{E}_z)}{\hbar^2}X \\ -\frac{\mathrm{d}^2 Z}{\mathrm{d}z^2} &= \frac{2m\mathcal{E}_z}{\hbar^2}Z. \end{aligned} \tag{11.14}$$

The physical content of these equations is clear: $\mathcal{E}_z$ is the kinetic energy associated with motion along the cylinder's axis, so motion perpendicular to the axis carries the remaining energy, $E_n - \mathcal{E}_z$. As we push in the piston, neither the equation governing X and $E_n - \mathcal{E}_z$ nor its boundary conditions change, so $E_n - \mathcal{E}_z$ is invariant. What does change is the boundary condition subject to which the equation for Z has to be solved.

We place one end of the cylinder at $z = 0$ and the piston at $z = L$. Then it is easy to see that the required solution for Z is [cf. §5.1.1(a)]

$$Z(z) \propto \sin(k\pi z/L) \quad \text{with} \quad k = 1, 2, \ldots, \tag{11.15}$$

and the possible values of $\mathcal{E}_z$ are

$$\mathcal{E}_z = \frac{h^2}{8mL^2}k^2. \tag{11.16}$$

The adiabatic principle assures us that if we let the piston out slowly, the particle's value of the quantum number k will not change, and its energy

$\mathcal{E}_z$ will evolve according to equation (11.16). By conservation of energy, the energy lost by the particle when L is increased by $\mathrm{d}L$ must equal the work that the particle does on the piston, which is $P\mathrm{d}\mathcal{V}$, where P is the pressure it exerts and $\mathrm{d}\mathcal{V}$ is the increase in the cylinder's volume. Let A be the area of the piston. Then conservation of energy requires that

$$-\mathrm{d}\mathcal{E}_z = 2\mathcal{E}_z\frac{\mathrm{d}L}{L} = PA\,\mathrm{d}L, \tag{11.17}$$

from which it follows that

$$P = \frac{2\mathcal{E}_z}{AL} = 2\frac{\mathcal{E}_z}{\mathcal{V}}. \tag{11.18}$$

When we sum the contributions to the pressure that arise from a large number, N, of molecules in the cylinder, equation (11.18) yields

$$P = 2\frac{N}{\mathcal{V}}\left\langle \mathcal{E}_z \right\rangle, \tag{11.19}$$

where the angle brackets mean the average over all molecules. At this point we have to take into account collisions between the N molecules. Colliding molecules change the directions of their momenta and thus transfer energy between motion in the z direction and motion in the plane perpendicular to it. Collisions do not satisfy the adiabatic approximation, so they do change the quantum numbers of particles. Their overall effect is to ensure that the velocity distribution remains isotropic even though the piston's motion is changing $\mathcal{E}_z$ and not the energy of motion in the plane of the piston, $E_n - \mathcal{E}_z$. So we may assume that $\langle E\rangle = 3\left\langle \mathcal{E}_z\right\rangle$. Let $U \equiv N\langle E\rangle$ be the internal energy of the gas. Then eliminating $\left\langle \mathcal{E}_z\right\rangle$ from equation (11.19) in favour of U, we obtain

$$P\mathcal{V} = \tfrac{2}{3}U. \tag{11.20}$$

This result is identical with what we obtain by combining the equation of state of an ideal gas, $P\mathcal{V} = Nk_{\mathrm{B}}T$, with the expression for the internal energy of such a gas, $U = \frac{3}{2}Nk_{\mathrm{B}}T$. Actually our result is more general than the result for an ideal gas because we have not assumed that the gas is in thermal equilibrium: the only assumption we have made about the distribution of kinetic energy among particles is that it is isotropic.

11.3 Application to thermodynamics

In §6.4 we saw that when a system is in thermodynamic equilibrium, we do not know what quantum state it is in but can assign a probability $p_i \propto \mathrm{e}^{-E_i/k_{\mathrm{B}}T}$ that it is in its i^{th} stationary state (eq. 6.93a). The energy E_i of this state depends on the variables, such as volume, electric field, shear stress, etc., that quantify the system's environment. In the simplest non-trivial case, that in which the system is a fluid, the only relevant variable is the volume $\mathcal{V}$ and we shall consider only this case. Hence we consider the energy of each stationary state to be a function $E_i(\mathcal{V})$.

In an adiabatic compression of our system, we slowly change $\mathcal{V}$ while isolating the system from heat sources. From the adiabatic principle it follows that during such a compression the system remains in whatever stationary state it was in when the compression started. Consequently, the probabilities p_i of its being in the various stationary states are constant, and the entropy $S = -k_{\mathrm{B}}\sum_i p_i \ln p_i$ (eq. 6.91) is constant during an adiabatic change, just as classical thermodynamics teaches.

During an adiabatic compression, the change in the internal energy $U = \sum_i p_i E_i$ is

$$\mathrm{d}U = \sum_i p_i \frac{\partial E_i}{\mathrm{d}\mathcal{V}} = -P\mathrm{d}\mathcal{V} \quad \text{where} \quad P \equiv -\sum_i p_i \frac{\partial E_i}{\partial \mathcal{V}}. \tag{11.21}$$

Since there is no heat flow, the increment in U must equal the work done, which is the pressure that the system exerts times $-\mathrm{d}\mathcal{V}$, so the quantity P defined by equation (11.21) is indeed the pressure.

11.4 The compressibility of condensed matter

As a second application of the adiabatic principle, we estimate the compressibility of solids and liquids. In condensed matter atoms touch one another and the volume of the bulk material can be reduced only if every atom is made smaller. If an atom's electrons are to be confined to a smaller volume, by the uncertainty principle, their momenta and therefore their kinetic energies must increase. We estimate the compressibility of matter by equating the work done in compressing it to the resulting increase in the energy of the atom. The adiabatic approximation tells us that during slow compression, the atom remains in its ground state. Hence the compressibility can be deduced if we can calculate the ground-state energy E_0 as a function of the atom's volume $\mathcal{V}$.

Compressibility χ is defined to be the fractional change in volume per unit applied pressure P:

$$\chi = -\frac{1}{\mathcal{V}}\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}P}. \tag{11.22}$$

Conservation of energy implies that $-P\,\mathrm{d}\mathcal{V}$, the work done by the compressor, is equal to the increase in the ground-state energy $\mathrm{d}E_0$, so $P = -\mathrm{d}E_0/\mathrm{d}\mathcal{V}$ and

$$\chi = \left(\mathcal{V}\frac{\mathrm{d}^2E_0}{\mathrm{d}\mathcal{V}^2}\right)^{-1}. \tag{11.23}$$

$E_0(\mathcal{V})$ can be obtained by solving for the atom's stationary states with the electronic wavefunction required to vanish on the surface of a sphere of volume $\mathcal{V}$. A highly simplified version of such a calculation enables us to obtain a rough estimate of the compressibility of condensed matter.

We assume that when the atom is confined in a sphere of radius a, its wavefunction $\langle\mathbf{x}|a\rangle$ is the same as the wavefunction for the confining sphere of radius a_0 with all distances rescaled by a/a_0 and the appropriate adjustment in the normalisation. In this case, we can argue as in §9.2 that the expectation value $\langle a|K|a\rangle$ of the atom's kinetic energy operator K scales as $(a_0/a)^2$, while the expectation value of the potential-energy operator V scales as a_0/a. Hence

$$\frac{\mathrm{d}E_0}{\mathrm{d}a} = \frac{\mathrm{d}}{\mathrm{d}a}\left(\langle a|K|a\rangle + \langle a|V|a\rangle\right) \simeq -2\frac{\langle a|K|a\rangle}{a} - \frac{\langle a|V|a\rangle}{a}. \tag{11.24}$$

Equation (8.48) states that $2\langle a|K|a\rangle = -\langle a|V|a\rangle$, so the right side of this equation vanishes.[1] Differentiating again, we find

$$\frac{\mathrm{d}^2E_0}{\mathrm{d}a^2} \simeq 6\frac{\langle a|K|a\rangle}{a^2} + 2\frac{\langle a|V|a\rangle}{a^2} = -2\frac{E_0}{a^2}, \tag{11.25}$$

where equation (8.48) has been used again to simplify the right side. Since $\mathcal{V} \propto a^3$, $\mathrm{d}\mathcal{V}/\mathrm{d}a = 3\mathcal{V}/a$ and bearing in mind our result that $\mathrm{d}E_0/\mathrm{d}a = 0$ we find

$$\frac{\mathrm{d}^2E_0}{\mathrm{d}\mathcal{V}^2} \simeq \left(\frac{a}{3\mathcal{V}}\right)^2\frac{\mathrm{d}^2E_0}{\mathrm{d}a^2} = -\tfrac{2}{9}\frac{E_0}{\mathcal{V}^2}. \tag{11.26}$$

Using this result in equation (11.23), we conclude that the compressibility is

$$\chi \simeq \tfrac{9}{2}\frac{\mathcal{V}}{|E_0|}. \tag{11.27}$$

Some care is required in the application of this result to many-electron atoms. Our assumption that $\langle a|V|a\rangle$ scales as a^{-1} is valid only if the wavefunction is simultaneously rescaled in the coordinates of *all* the atom's electrons. Unfortunately, it is physically obvious that, at least for small fractional

---

[1] Equation (8.48) was actually only derived for hydrogen, but the result applies to the gross structure of any atom.

changes in volume, only the outermost shell of electrons will be significantly affected by the confining sphere. So realistically we should assume that the system formed by the inner electron shells remains fixed and the wavefunction is rescaled in its dependence on the coordinates of electrons in the outermost shell. In this spirit we shall replace $|E_0|$ by N times the atom's ionisation energy, where N is the number of electrons in the outermost shell. Since the electrostatic potential produced by the nucleus and the fixed inner shells of electrons does not vary with radius as r^{-1}, $\langle a|V|a\rangle$ will not scale as a^{-1} and the factor $\frac{9}{2}$ in equation (11.27) will be in error. None the less, the equation should correctly predict the order of magnitude of an atom's compressibility.

For lithium we take $\mathcal{V} = \frac{4}{3}\pi(2a_0)^3$ and $|E_0| = 5.39\,\mathrm{eV}$ to find $\chi = 2.6\times10^{-11}\,\mathrm{Pa}^{-1}$. The measured value varies with temperature and is of order $10^{-10}\,\mathrm{Pa}^{-1}$, which is in excellent agreement with our quantum-mechanical estimate given the sensitivity of the latter to the adopted value of the rather ill-defined parameter $\mathcal{V}$.

11.5 Covalent bonding

The air we breathe, the living tissue of our bodies, and the plastics in the clothes, chairs and carpets that surround us, are held together by **covalent bonds**. These are bonds between two atoms of similar electronegativity, such as two oxygen atoms, two carbon atoms or a carbon atom and an oxygen atom. In this section we explain how they arise through the sharing by the atoms of one or more electrons. Unlike the ionic bonds that hold together a crystal of common salt, which are crudely electrostatic in nature, a covalent bond is intrinsically quantum-mechanical.

11.5.1 A toy model of a covalent bond

To show how covalent bonding works, we study a one-dimensional toy model that is not at all realistic but it is analytically tractable. We imagine a particle of mass m that moves along the x axis in a potential $V(x)$ that is made up of two δ-function potentials of the type we introduced in §5.1.1(b). The wells are separated by distance $2a$:

$$V(x) = -V_\delta\{\delta(x+a) + \delta(x-a)\}. \tag{11.28}$$

We have placed the origin at the midpoint between the two wells, which we can do without loss of generality. This placement ensures that the Hamiltonian commutes with the parity operator, and we can seek solutions of the TISE that have well-defined parity. There are three distinct regions in which $V(x) = 0$, namely $x < -a$, $-a < x < a$ and $x > a$ and in these regions the wavefunction $u(x)$ must be a linear combination of the exponentials $\mathrm{e}^{\pm kx}$, where k is related to E by

$$k = \sqrt{-2mE}\,/\hbar. \tag{11.29}$$

With an eye to the construction of solutions of definite parity we let our solutions in these regions be

$$u(x) \propto \begin{cases} \mathrm{e}^{kx} & \text{for } x < -a, \\ \cosh(kx) \text{ or } \sinh(kx) & \text{for } -a < x < a, \\ \mathrm{e}^{-kx} & \text{for } x > a, \end{cases} \tag{11.30}$$

Bearing in mind that the wavefunction has to be continuous across the potential wells, we see from Figure 11.2 that solutions of each parity must be

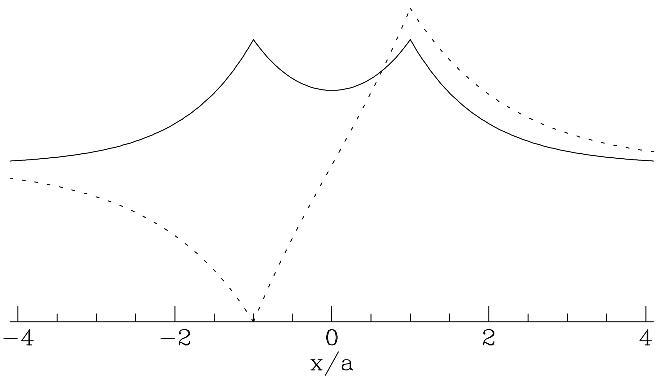


Figure 11.2 Even- and odd-parity solutions to the two delta-function problem.

of the form

$$\begin{aligned} u_+(x) &= A \times \begin{cases} \mathrm{e}^{k(x+a)} & \text{for } x < -a, \\ \cosh(kx)/\cosh(ka) & \text{for } -a < x < a, \\ \mathrm{e}^{-k(x-a)} & \text{for } x > a, \end{cases} \\ u_-(x) &= B \times \begin{cases} -\mathrm{e}^{k(x+a)} & \text{for } x < -a, \\ \sinh(kx)/\sinh(ka) & \text{for } -a < x < a, \\ \mathrm{e}^{-k(x-a)} & \text{for } x > a, \end{cases} \end{aligned} \tag{11.31}$$

where the constants on each segment of the real line have been chosen to ensure that the wavefunctions equal A and B, respectively, at $x = a$.

On account of the symmetry of the problem, it suffices to choose k for each parity such that the equation (5.13) is satisfied at $x = a$. From this equation we have that

$$2K = \begin{cases} k\{1 + \tanh(ka)\} & \text{for even parity} \\ k\{1 + \coth(ka)\} & \text{for odd parity} \end{cases} \tag{11.32}$$

where K is defined by equation (5.14). By expressing the hyperbolic functions in terms of e^{ka}, we can rearrange the equations into

$$\frac{k}{K} - 1 = \pm \mathrm{e}^{-2ka}, \tag{11.33}$$

where the upper sign is for even parity. In the upper panel of Figure 11.3 the left and right sides of these equations are plotted; the solution k is the ordinate at which the straight line of the left side intersects with the decaying exponential plot of the right side. The value of k_+ that we obtain from the upper curve associated with the even-parity case is always larger than the value k_- obtained for the odd-parity case. By equation (11.29), the particle's binding energy increases with k, so the even-parity state is the more tightly bound. If we increase a, the exponential curves in the top panel of Figure 11.3 become more curved and approach the x-axis more rapidly. Hence k_+ diminishes, and k_- grows. In the limit $a \to \infty$, the exponentials hug the axes ever more tightly and k_+ and k_- converge on the point $k = K$ at which the sloping line crosses the x-axis. This value of k is precisely that for an isolated well as we would expect, since in the limit $a \to \infty$ the wells are isolated from one another. The lower panel of Figure 11.3 shows the energies associated with $k_\pm$ from equation (11.29).

Suppose our particle is initially in the ground state of two wells that are distance $2a$ apart, and imagine slowly moving the two wells towards one another. By the adiabatic principle, the particle stays in the ground state, which, as we have seen, moves to lower energies. Hence the particle loses energy. Where does this energy go? It can only go into the mechanism that

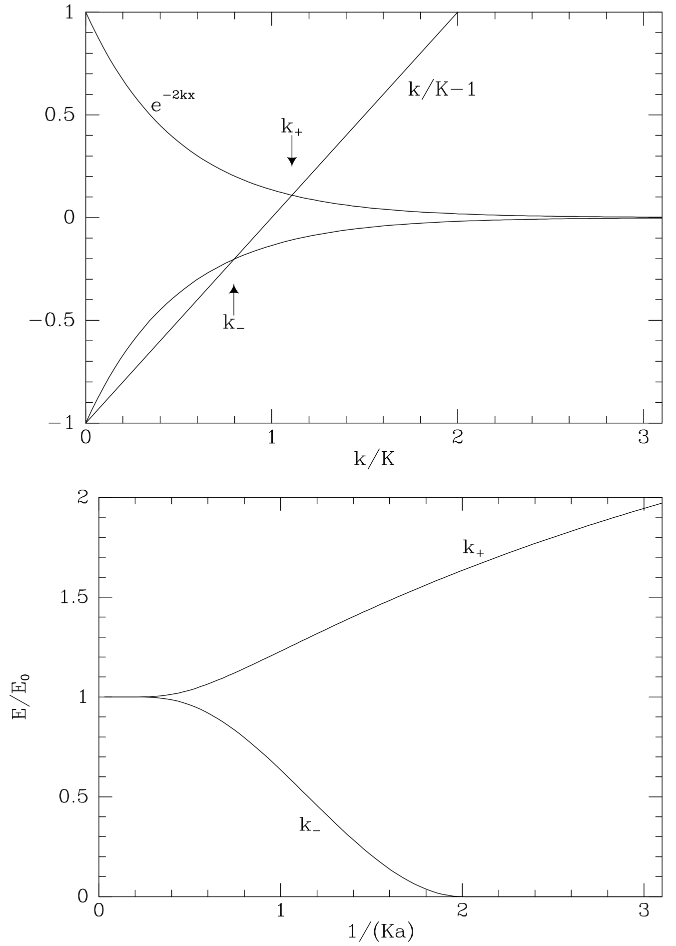


Figure 11.3 Graphical solution of equation (11.33). In the top panel, the exponential is drawn for the case $Ka = 1$. Bottom panel: binding energy versus inverse separation. The scale energy $E_0 = -\hbar^2K^2/2m$ is the energy of the bound state of an isolated δ-function potential well.

positions the wells. A little thought reveals that if work is done on this mechanism, it must be resisting the mutual attraction by the holes. Hence we have arrived at the conclusion that two potential wells that are jointly binding a particle, can experience a mutual attraction that would not be present if the bound particle were absent.

11.5.2 Molecular dynamics

The toy model just presented describes an inherently quantum-mechanical mechanism by which atoms experience mutual attraction or repulsion through sharing a bound particle. An essentially identical calculation, in which the energy E_e of the two shared electrons on an H_2 molecule is studied as a function of the separation b of the protons, enables one to understand the structure of the H_2 molecule. Analogously with the toy model, the energy of the shared electrons decreases monotonically with b, so in the absence of the mutual electrostatic repulsion of the protons, which had no analogue in our model, the electrons would draw the protons ever closer together. In reality there is a particular value b_0 of b at which the rate at which $E_\mathrm{e}(b)$ decreases equals the rate at which the electrostatic energy $E_\mathrm{p}(b)$ of the protons increases with decreasing b. The classical intranuclear separation of an H_2 molecule is b_0.

A more complete theory is obtained by considering $V(b) \equiv E_\mathrm{e} + E_\mathrm{p}$ to be a dynamical potential in which the two nuclei move. The analysis of this problem proceeds in close analogy with our treatment of the hydrogen atom in §8.1: we introduce a reduced particle and observe that the Hamiltonian that governs its dynamics commutes with the reduced particle's angular-momentum operator; this observation enables us to derive a radial wave equation for each value of the angular-momentum quantum number l. This radial wave equation describes oscillations that are governed by the **effective potential** $V(b)$. The rotation-vibration spectrum of H_2 may be understood as arising from transitions between states that are characterised by l and the quantum number n associated with the oscillations in b.

Similar principles clearly apply to studies of the dynamics of many other molecules: one starts by determining the energy of shared electrons for a grid of fixed locations of the nuclei. The resulting energies together with the energy of electrostatic repulsion between the nuclei yields an effective potential, that can then be used to study the quantum dynamics of the nuclei. The essential approximation upon which this kind of work depends is that the frequencies at which the nuclei oscillate are low compared to any *difference* between energy levels of the electronic system, divided by $\hbar$. Since electrons are so much lighter than nuclei, this approximation is generally an excellent one when the molecular rotations and vibrations are not strongly excited. The approximation is guaranteed to break down during dissociation of a molecule, however. We return to the toy model to explain why this is so.

11.5.3 Dissociation of molecules

In the model of §11.5.1, the force provided by the particle that is shared between the potential wells is not always attractive: if the particle in the excited, odd-parity bound-state the energy of the particle *increases* as the separation of the wells $2a$ is diminished, so the positioning mechanism must be pushing the two wells together as it resists the mutual repulsion of the wells. Consider now a two-well molecule that is held together by the attractive force provided by the shared particle in its ground state when a photon promotes the particle to its excited state. Then the force provided by the particle becomes repulsive, and the wells will begin to move apart. As they move, much of the energy stored in the excitation of the particle is converted into kinetic energy of the wells, and soon there is one bare well and one well with a trapped particle.

As the wells move apart, the energy difference between the ground and excited states decreases, while the rate of increase of a increases. Hence the adiabatic approximation, which requires that $a/\dot{a} \ll (E_o - E_e)/\hbar$ must break down. In a more complex system, such as a real CO molecule, this breakdown can cause some of the energy stored in the particle's excitation being transferred to excitation of one or both of the final atoms rather than being converted to the kinetic energy of their motion apart.

11.6 The WKBJ approximation

In §5.4 we learnt from a numerical solution of the TISE that when a particle encounters a modest ramp in its potential energy V, the amplitude for reflection is small unless the distance over which V changes is small compared to the particle's de Broglie wavelength. This result is closely related to the adiabatic approximation in the sense that in the particle's rest frame, the potential that it experiences changes slowly compared to the time taken to cover a de Broglie wavelength. Now we given an important analytical argument that leads to the same conclusion and allows us to determine the evolution of the wave as it moves over the barrier.

The equation we wish to solve is

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = -k^2\psi \quad \text{where} \quad k^2(x) \equiv \frac{2m}{\hbar^2}(E - V), \tag{11.34}$$

and $V(x)$ is some real function. We define

$$\phi(x) \equiv \int^x \mathrm{d}x'\, k(x') = \frac{\sqrt{2m}}{\hbar}\int^x \mathrm{d}x'\, \sqrt{E - V(x')}, \tag{11.35}$$

so $k = \mathrm{d}\phi/\mathrm{d}x$. Then without loss of generality we write

$$\psi(x) = S(x)\mathrm{e}^{\mathrm{i}\phi}, \tag{11.36}$$

where S is a function to be determined. When we substitute this expression for ψ into the TISE (11.34), cancel the right side with one of the terms on the left and then divide through by $\mathrm{e}^{\mathrm{i}\phi}$, we obtain

$$\frac{\mathrm{d}^2 S}{\mathrm{d}x^2} + 2\mathrm{i}k\frac{\mathrm{d}S}{\mathrm{d}x} + \mathrm{i}S\frac{\mathrm{d}k}{\mathrm{d}x} = 0. \tag{11.37}$$

Now we reason that when k is a slowly varying function of x, S will be also. In fact, if k changes on a lengthscale $L \gg 1/k$ that is greater than the wavelength $2\pi/k$, we will have $|\mathrm{d}k/\mathrm{d}x| \sim |k|/L$, $|\mathrm{d}S/\mathrm{d}x| \sim |S|/L$ and $|\mathrm{d}^2S/\mathrm{d}x^2| \sim |S|/L^2$. In these circumstances it follows that the first term in equation (11.37) is negligible compared to the other two, and we may approximate the equation by

$$\frac{\mathrm{d}\ln S}{\mathrm{d}x} = -\tfrac{1}{2}\frac{\mathrm{d}\ln k}{\mathrm{d}x} \tag{11.38}$$

Integrating both sides from x_1 to x_2 we find that

$$(S\sqrt{k})\big|_{x_1} = (S\sqrt{k})\big|_{x_2}. \tag{11.39}$$

The particle flux implied by $\psi(x)$ is proportional to the probability density $|S|^2$ times the particle speed k/m. Hence equation (11.39) states that the particle flux at x_1 is equal to that at x_2. In other words, *when the wavenumber changes very little in a wavelength, the reflected amplitude is negligible and the wavefunction is approximately*

$$\psi(x) \simeq \text{constant} \times \left(\frac{\hbar^2}{2m(E - V)}\right)^{1/4} \mathrm{e}^{\mathrm{i}\phi} \tag{11.40}$$

*where $\phi(x)$ is given by equation (*11.35*).* This solution is known as the **WKBJ approximation**.[2] The WKBJ approximation guarantees conservation of particle flux in the classical limit of very small de Broglie wavelengths. It also has innumerable applications outside quantum mechanics, including the working of ear trumpets, tidal bores and Saturn's rings.

---

[2] The WKBJ approximation is named after Wentzel, Kramers, Brillouin and Jeffreys. This is frequently abbreviated to 'WKB approximation'.

Problems

11.1 In §9.1 we obtained estimates of the amount by which the energy of an atom changes when an electric or magnetic field is applied. Discuss whether the derivation of these results implicitly assumed the validity of the adiabatic principle.

11.2 In §11.2 we assumed that the potential energy of air molecules is infinitely large inside a bicycle pump's walls. This cannot be strictly true. Give a reasoned order-of-magnitude estimate for the potential in the walls, and consider how valid it is to approximate this by infinity.

11.3 Explain why E/ω is an adiabatic invariant of a simple harmonic oscillator, where ω is the oscillator's angular frequency. Einstein proved this result in classical physics when he was developing the "old quantum theory", which involved quantising adiabatic invariants such as E/ω and angular momentum. Derive the result for a classical oscillator by adapting the derivation of the WKBJ approximation to the oscillator's equation of motion $\ddot{x} = -\omega^2 x$.

11.4 Consider a particle that is trapped in a one-dimensional potential well $V(x)$. If the particle is in a sufficiently highly excited state of this well, its typical de Broglie wavelength may be sufficiently smaller than the characteristic lengthscale of the well for the WKBJ approximation to be valid. Explain why it is plausible that in this case

$$\frac{1}{\hbar}\int_{x_1}^{x_2} \mathrm{d}x' \sqrt{2m\{E - V(x')\}} = n\pi, \tag{11.41}$$

where the $E - V(x_i) = 0$ and n is an integer. Relate this condition to the quantisation rule $\oint \mathrm{d}x\, p_x = nh$ used in the "old quantum theory".

11.5 Show that the "old quantum theory" (Problem 11.4) predicts that the energy levels of the harmonic oscillator are $n\hbar\omega$ rather than $(n+\frac{1}{2})\hbar\omega$. Comment on the dependence on n of the fractional error in E_n.

11.6 Suppose the charge carried by a proton gradually decayed from its current value, e, being at a general time fe. Write down an expression for the binding energy of a hydrogen atom in terms of f. As $\alpha \to 0$ the binding energy vanishes. Explain physically where the energy required to free the electron has come from.

When the spring constant of an oscillator is adiabatically weakened by a factor f^4, the oscillator's energy reduces by a factor f^2. Where has the energy gone?

In Problems 3.12 and 3.13 we considered an oscillator in its ground state when the spring constant was suddenly weakened by a factor $f = 1/16$. We found that the energy decreased from $\frac{1}{2}\hbar\omega$ to $0.2656\hbar\omega$ not to $\hbar\omega/512$. Explain physically the difference between the sudden and adiabatic cases.

11.7 Photons are trapped inside a cavity that has perfectly reflecting walls which slowly recede, increasing the cavity's volume $\mathcal{V}$. Give a physical motivation for the assumption that each photon's frequency $\nu \propto \mathcal{V}^{-1/3}$. Using this assumption, show that the energy density of photons $u \propto \mathcal{V}^{-4/3}$ and hence determine the scaling with $\mathcal{V}$ of the pressure exerted by the photons on the container's walls.

Black-body radiation comprises an infinite set of thermally excited harmonic oscillators – each normal mode of a large cavity corresponds to a new oscillator. Initially the cavity is filled with black-body radiation of temperature T_0. Show that as the cavity expands, the radiation continues to be black-body radiation although its temperature falls as $\mathcal{V}^{-1/3}$. Hint: use equation (6.118).

11.8 Show that when a charged particle gyrates with energy E in a uniform magnetic field of flux density B, the **magnetic moment** $\mu = E/B$ is invariant when B is changed slowly. Hint: recall Problem 3.22. By applying the principle that energy must be conserved when the magnetic field is slowly ramped up, deduce whether a plasma of free electrons forms a para- or dia-magnetic medium.

12

Scattering Theory

Up to this point our discussion of the dynamics of particles has concentrated on situations in which a particle is trapped in a potential well, with the consequence that its wavefunction tends to zero at infinity. In reality we often have to deal with situations in which a free particle approaches a region of enhanced potential from infinity, is deflected, and moves away in a new direction. For example, we see by catching photons that have been scattered in this way as they travel out from a light source.

All around us molecules in the air are scattering off one another. This scattering is sometimes inelastic and leaves a molecule in an excited state that subsequently decays with the emission of radiation of a characteristic frequency. Hence we have to understand these unseen scattering experiments if we want to understand the spectral distribution of the electromagnetic radiation emitted by air.

Most of what we know about the small-scale structure of matter has been gleaned from carefully monitored scattering experiments. Rutherford showed that atoms have compact nuclei by studying the distribution of α particles scattered by gold foil, and we now study the sub-atomic structure of matter by scattering extremely fast-moving electrons and protons off one another in high-energy accelerators. The theoretical tools required by such experiments are rather different from those used to determine the spectral distribution of radiation from a gas because in the latter case, as in our studies of the harmonic oscillator and the hydrogen atom, we know the potential $V(\mathbf{x})$ that controls the dynamics. By contrast, when we use scattering experiments to probe the small-scale structure of matter we wish to infer the unknown potential from the observed dynamics. Much of scattering theory is directed towards problems of this kind.

12.1 The scattering operator

Let $|\psi\rangle$ be the state of a particle in a scattering experiment. The evolution of $|\psi\rangle$ is governed by the TDSE

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle \quad \Leftrightarrow \quad |\psi;t\rangle = U(t)|\psi;0\rangle, \tag{12.1a}$$

where $U(t) = \mathrm{e}^{-\mathrm{i}Ht/\hbar}$ is the time evolution operator introduced in §4.3. We break the Hamiltonian into a sum $H = H_\mathrm{K}+V$ of the kinetic-energy operator $H_\mathrm{K} = p^2/2m$ and the potential V that causes scattering. If $|\psi\rangle$ represents a moving particle – one that approaches or leaves some collision – it must be a non-trivial superposition of the eigenstates of H – see §2.3.3. Unfortunately, when $V \neq 0$ we may not know what these eigenstates are, and it may be prohibitively difficult to find them.

A crucial physical insight allows us to make progress: in a scattering experiment, long before the particle reaches the interaction region it is approximately free. The evolution of a free particle $|\phi\rangle$ is governed by H_K

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\phi\rangle = H_\mathrm{K}|\phi\rangle \quad \text{or} \quad |\phi;t\rangle = U_\mathrm{K}(t)|\phi;0\rangle, \tag{12.1b}$$

where $U_\mathrm{K}(t) = \mathrm{e}^{-\mathrm{i}H_\mathrm{K}t/\hbar}$, so the statement that $|\psi\rangle$ behaves like a free particle in the asymptotic past is the requirement that

$$\lim_{t\to-\infty} U(t)|\psi;0\rangle \ = \lim_{t\to-\infty} U_\mathrm{K}(t)|\phi;0\rangle \tag{12.2a}$$

for some free state $|\phi\rangle$. Implicit in this equation is the assumption that the origin of time is chosen such that the interaction takes place at some finite time. For example, it might be in full swing at time $t = 0$, which we shall sometimes refer to as 'the present'.

In the asymptotic future the scattered particle will have moved far away from the interaction region and will again be approximately free. Hence we also require

$$\lim_{t\to+\infty} U(t)|\psi;0\rangle \ = \lim_{t\to+\infty} U_\mathrm{K}(t)|\phi';0\rangle \tag{12.2b}$$

where $|\phi'\rangle$ is another free state.

Equations (12.2) allow us to relate the real state $|\psi\rangle$ to both $|\phi\rangle$ and $|\phi'\rangle$ at the present time via

$$\begin{aligned}|\psi;0\rangle \ &= \lim_{t\to-\infty} U^\dagger(t)U_\mathrm{K}(t)|\phi;0\rangle \ = \lim_{t'\to+\infty} U^\dagger(t')U_\mathrm{K}(t')|\phi';0\rangle \\ &= \Omega_+|\phi;0\rangle = \Omega_-|\phi';0\rangle,\end{aligned} \tag{12.3}$$

where the operators $\Omega_\pm$ are defined by[1]

$$\Omega_\pm \ \equiv \lim_{t\to\mp\infty} U^\dagger(t)U_\mathrm{K}(t). \tag{12.4}$$

The origin of the irritating choices of sign in this definition will be explained in §12.2 below. In terms of the $\Omega_\pm$ operators, the **scattering operator** $\mathcal{S}$ is defined as

$$\mathcal{S} \equiv \Omega_-^\dagger\Omega_+. \tag{12.5}$$

Here's what the scattering operator does: first, $\mathcal{S}$ evolves a free-particle state back to the distant past, then matches it onto a real state which has the same past asymptotic behaviour. Next, $\mathcal{S}$ evolves this real state forwards – all the way through the scattering process to the far future. There, the real particle again behaves like a free particle, and $\mathcal{S}$ matches their states before finally

---

[1] It is not self-evident that the limits as $t \to \pm\infty$ that appear in equation (12.4) exist. Appendix H derives a condition on V that ensures that $\Omega_\pm$ is well defined.

evolving the free state back to the present. If the real particle is in some state $|\psi\rangle$, and looks like a free-particle state $|\phi\rangle$ well before the interaction occurs, then the amplitude for it will look like some other free state $|\lambda\rangle$ long after the interaction is $\langle\lambda|\mathcal{S}|\phi\rangle$. Hence the probability to find the particle in the free state $|\lambda\rangle$ is just $|\langle\lambda|\mathcal{S}|\phi\rangle|^2$. The scattering operator is useful because it always acts on *free* states, so if we use it we do not need to know the eigenstates of the full Hamiltonian H.

Notice that $\mathcal{S}$ is defined as a product of four unitary evolution operators and is therefore itself unitary.

When $V = 0$, the particle isn't scattered, and its future state is the same as its past state. In such circumstances, the scattering operator must be just the identity operator $S = 1$, and we can check this is indeed true by putting $H = H_{\mathrm{K}}$ in equations (12.3) to (12.5). The operator that describes a genuine interaction is the **transition operator**

$$\mathcal{T} \equiv \mathcal{S} - 1 \tag{12.6}$$

and the probabilities for actual transitions are given by

$$\mathrm{Prob}(|\phi\rangle \to |\lambda\rangle) = |\langle\lambda|\mathcal{T}|\phi\rangle|^2 = |\langle\lambda|\mathcal{S}|\phi\rangle - \langle\lambda|\phi\rangle|^2 . \tag{12.7}$$

Since $\mathcal{S}$ is unitary, we have

$$1 = \mathcal{S}^\dagger\mathcal{S} = 1 + \mathcal{T}^\dagger + \mathcal{T} + \mathcal{T}^\dagger\mathcal{T}. \tag{12.8}$$

Squeezing this equation between $\langle\phi|$ and $|\phi\rangle$ we obtain

$$-2\Re\mathrm{e}\{\langle\phi|\mathcal{T}|\phi\rangle\} = \langle\phi|\mathcal{T}^\dagger\mathcal{T}|\phi\rangle. \tag{12.9}$$

We also have that

$$|\langle\phi|\mathcal{S}|\phi\rangle|^2 = |1 + \langle\phi|\mathcal{T}|\phi\rangle|^2 = 1 + 2\Re\mathrm{e}\{\langle\phi|\mathcal{T}|\phi\rangle\} + |\langle\phi|\mathcal{T}|\phi\rangle|^2. \tag{12.10}$$

Rearranging and using equation (12.9) yields

$$1 - |\langle\phi|\mathcal{S}|\phi\rangle|^2 = \sum_{|\psi_i\rangle\neq|\phi\rangle} \langle\phi|\mathcal{T}^\dagger|\psi_i\rangle\langle\psi_i|\mathcal{T}|\phi\rangle = \sum_{|\psi_i\rangle\neq|\phi\rangle} |\langle\psi_i|\mathcal{T}|\phi\rangle|^2 , \tag{12.11}$$

where $\{|\psi_i\rangle\}$ is a complete set of states that includes $|\phi\rangle$. The left side of equation (12.11) is one, minus the probability that at $t = +\infty$ the particle is still in the state it was in at $t = -\infty$, while the right side is the sum of the probabilities that the particle has made the transition to some state $|\psi_i\rangle$ different from the original state $|\phi\rangle$.

12.1.1 Perturbative treatment of the scattering operator

The definition $\mathcal{S} \equiv \Omega_-^\dagger\Omega_+$ is difficult to use in practical calculations, because the true Hamiltonian H (whose eigenstates we do not know) is buried rather deep inside. To get at it we first differentiate $\Omega(t) \equiv U^\dagger(t)U_{\mathrm{K}}(t)$ with respect to t, finding

$$\frac{\mathrm{d}}{\mathrm{d}t}\Omega(t) = \frac{\mathrm{i}}{\hbar}\mathrm{e}^{\mathrm{i}Ht/\hbar}(H - H_{\mathrm{K}})\mathrm{e}^{-\mathrm{i}H_{\mathrm{K}}t/\hbar} = \frac{\mathrm{i}}{\hbar}\mathrm{e}^{\mathrm{i}Ht/\hbar}V\mathrm{e}^{-\mathrm{i}H_{\mathrm{K}}t/\hbar}, \tag{12.12}$$

where we have been careful to preserve the order of the operators. We now re-integrate this equation between t' and t to reach

$$\begin{aligned}\Omega(t) &= \Omega(t') + \frac{\mathrm{i}}{\hbar}\int_{t'}^{t}\mathrm{d}\tau\,\mathrm{e}^{\mathrm{i}H\tau/\hbar}V\mathrm{e}^{-\mathrm{i}H_{\mathrm{K}}\tau/\hbar}\\ &= \Omega(t') + \frac{\mathrm{i}}{\hbar}\int_{t'}^{t}\mathrm{d}\tau\,U^\dagger(\tau)VU_{\mathrm{K}}(\tau).\end{aligned} \tag{12.13}$$

Taking the Hermitian adjoint of this equation we have

$$\begin{aligned}\Omega^\dagger(t) &= \Omega^\dagger(t') - \frac{\mathrm{i}}{\hbar}\int_{t'}^{t}\mathrm{d}\tau\, U_{\mathrm{K}}^\dagger(\tau)VU(\tau)\\ &= \Omega^\dagger(t') - \frac{\mathrm{i}}{\hbar}\int_{t'}^{t}\mathrm{d}\tau\, U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau)\Omega^\dagger(\tau).\end{aligned} \tag{12.14}$$

The integrand itself contains $\Omega^\dagger(\tau)$, but suppose we use this equation to replace it by $\Omega(t')$ plus an integral that involves $\Omega^\dagger(\tau')$, and then repeat this process once more. The result is

$$\begin{aligned}\Omega^\dagger(t) = \Omega^\dagger(t') &- \frac{\mathrm{i}}{\hbar}\int_{t'}^{t}\mathrm{d}\tau U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau)\,\Omega^\dagger(t')\\ &- \frac{1}{\hbar^2}\int_{t'}^{t}\mathrm{d}\tau\int_{t'}^{\tau}\mathrm{d}\tau'\, U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau-\tau')VU_{\mathrm{K}}(\tau')\Omega^\dagger(t')\\ &+ \frac{\mathrm{i}}{\hbar^3}\int_{t'}^{t}\mathrm{d}\tau\int_{t'}^{\tau}\mathrm{d}\tau'\int_{t''}^{\tau}\mathrm{d}\tau''\, U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau-\tau')VU_{\mathrm{K}}(\tau'-\tau'')VU_{\mathrm{K}}(\tau'')\Omega^\dagger(\tau'').\end{aligned} \tag{12.15}$$

Through repeated use of equation (12.14) we can push the operator $\Omega^\dagger(\tau)$ for $\tau > t'$ off into an integral that contains as many powers of V as we please. For sufficiently small V, the magnitude of the term in which $\Omega(\tau)$ occurs will be negligible, and we will be able to drop it. Then multiplying the equation by $\Omega(t')$, and taking the limits $t \to \infty$ and $t' \to -\infty$, we obtain an expansion of $\mathcal{S}$ in powers of V. Since $\Omega^\dagger(t')\Omega(t') = 1$, this expansion is

$$\begin{aligned}\mathcal{S} = 1 &- \frac{\mathrm{i}}{\hbar}\int_{-\infty}^{\infty}\mathrm{d}\tau\; U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau)\\ &- \frac{1}{\hbar^2}\int_{-\infty}^{\infty}\mathrm{d}\tau\int_{-\infty}^{\tau}\mathrm{d}\tau'\; U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau-\tau')VU_{\mathrm{K}}(\tau')\\ &+ \frac{\mathrm{i}}{\hbar^3}\int_{-\infty}^{\infty}\mathrm{d}\tau\int_{-\infty}^{\tau}\mathrm{d}\tau'\int_{-\infty}^{\tau'}\mathrm{d}\tau''\; U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau-\tau')VU_{\mathrm{K}}(\tau'-\tau'')VU_{\mathrm{K}}(\tau'')\\ &+ \cdots\end{aligned} \tag{12.16}$$

The virtue of equation (12.16) is that all the evolution operators involve only the free Hamiltonian H_{K} – information about scattering has been encoded in the expansion in powers of V.[2]

Equation (12.16) has an intuitive physical interpretation. The zeroth-order term is the identity operator and represents no scattering; its presence was anticipated by equation (12.6). The term $\mathcal{S}^{(1)}$ with one power of V acts on a free particle as

$$\begin{aligned}\langle\lambda;0|\mathcal{S}^{(1)}|\phi;0\rangle &= -\frac{\mathrm{i}}{\hbar}\int_{-\infty}^{\infty}\mathrm{d}\tau\langle\lambda;0|U_{\mathrm{K}}^\dagger(\tau)VU_{\mathrm{K}}(\tau)|\phi;0\rangle\\ &= -\frac{\mathrm{i}}{\hbar}\int_{-\infty}^{\infty}\mathrm{d}\tau\langle\lambda;\tau|V|\phi;\tau\rangle.\end{aligned} \tag{12.17}$$

The integrand $\langle\lambda;\tau|V|\phi;\tau\rangle$ is the amplitude for a particle in the free state $|\phi\rangle$ to be deflected by the potential V at time τ, transferring it into another free state $|\lambda\rangle$. Since we only observe the initial and final states, we do not know when the interaction took place, so we add the amplitudes for the deflection

---

[2] This expansion is reminiscent of the perturbation theory developed in §9.1. However, that theory hinged on the assumption that the response of the system to changes in its Hamiltonian is analytic in the parameter β. Here we need no such assumption. Instead we guess that certain integrals become small.

to have occurred at any time. Similarly, the second-order term $\mathcal{S}^{(2)}$ gives the amplitude

$$\langle\lambda;0|\mathcal{S}^{(2)}|\phi;0\rangle = \left(\frac{\mathrm{i}}{\hbar}\right)^2 \int_{-\infty}^{\infty}\mathrm{d}\tau \int_{-\infty}^{\tau}\mathrm{d}\tau'\ \langle\lambda;\tau|VU_\mathrm{K}(\tau-\tau')V|\phi;\tau'\rangle \tag{12.18}$$

for an incoming particle in the free state $|\phi\rangle$ to be deflected by the potential at time τ', then to propagate freely for a further time $\tau-\tau'$, and finally to be deflected again by V into the final state $|\lambda\rangle$. Since we do not know when either deflection occurred, we integrate over all τ and τ', subject to the condition $\tau' < \tau$ that the first deflection happens earlier. Higher-order terms describe trajectories that involve larger numbers of deflections.

If the potential is sufficiently weak, we might hope to approximate equation (12.16) by its lowest-order terms

$$\mathcal{S} \simeq 1 - \frac{\mathrm{i}}{\hbar}\int_{-\infty}^{\infty}\mathrm{d}\tau\ U_0^\dagger(\tau)VU_\mathrm{K}(\tau). \tag{12.19}$$

This drastic curtailing of the series for $\mathcal{S}$ is known as the **Born approximation**. Whether it is a good approximation in a given physical situation must be checked, often by estimating the order-of-magnitude of the second-order term $\mathcal{S}^{(2)}$, and checking it is acceptably smaller than the Born term.

12.2 The S-matrix

It is impossible to put any physical particle into a pure energy eigenstate, because such states are not localised in time. Nonetheless, energy eigenstates are useful as mathematical tools, being simpler to handle than realistic superpositions. Calculating with energy eigenstates presents special problems in scattering theory, because the idea that the particle *moves* towards the potential is central to our entire formalism, but a particle that is in an energy eigenstate goes nowhere.

12.2.1 The iϵ prescription

To get to the root of the problem, notice that equation (12.4) implies that

$$\begin{aligned} H\Omega_\pm &= \mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathrm{e}^{-\mathrm{i}Ht/\hbar}\,\Omega_\pm \\ &= \mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\left(\lim_{\tau\to\mp\infty}\mathrm{e}^{-\mathrm{i}Ht/\hbar}U^\dagger(\tau)U_\mathrm{K}(\tau)\right) \\ &= \mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\left(\lim_{\tau\to\mp\infty}U^\dagger(\tau-t)U_\mathrm{K}(\tau-t)U_\mathrm{K}(t)\right) \\ &= \mathrm{i}\hbar\Omega_\pm\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\mathrm{e}^{-\mathrm{i}H_\mathrm{K}t/\hbar} = \Omega_\pm H_\mathrm{K}. \end{aligned} \tag{12.20}$$

Therefore, if the interacting state $|\psi\rangle$ initially resembles some eigenstate $|E;\mathrm{free}\rangle$ of the free Hamiltonian, then equations (12.2a) and (12.20) imply that

$$H|\psi\rangle = H\Omega_+|E;\mathrm{free}\rangle = \Omega_+H_\mathrm{K}|E;\mathrm{free}\rangle = E\Omega_+|E;\mathrm{free}\rangle = E|\psi\rangle, \tag{12.21}$$

so $|\psi\rangle$ must actually be an eigenstate $|E;\mathrm{true}\rangle$ of the true Hamiltonian, with the same energy E. The trouble with this is that energy eigenstates look the same at all times, so $|E;\mathrm{true}\rangle$ and $|E;\mathrm{free}\rangle$ would always look like each other – a state of affairs that only makes sense if $V=0$ and there is no scattering.

Our argument shows that the initial and final states cannot have well-defined energy – they must be non-trivial superpositions of energy eigenstates. Nonetheless, because energy eigenstates often simplify otherwise difficult calculations, we are reluctant to forego them. Instead, we seek a way to avoid the problem. From equation (12.13), write $\Omega_\pm$ in the form

$$\begin{aligned}\Omega_\pm &= 1 + \frac{\mathrm{i}}{\hbar}\int_0^{\mp\infty} \mathrm{d}\tau\, U^\dagger(\tau) V U_{\mathrm{K}}(\tau) \\ &= 1 + \frac{\mathrm{i}}{\hbar}\int_0^{\mp\infty} \mathrm{d}\tau \int \mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{x}'\, U^\dagger(\tau)|\mathbf{x}\rangle\langle\mathbf{x}|V|\mathbf{x}'\rangle\langle\mathbf{x}'|U_{\mathrm{K}}(\tau)\ .\end{aligned} \tag{12.22}$$

When $\Omega_\pm$ is applied to a real scattering state $|\phi\rangle$, the integrand vanishes for large τ, because $\langle\mathbf{x}|U_{\mathrm{K}}|\phi\rangle$ is non-negligible only far from the scattering centre, where $\langle\mathbf{x}|V|\mathbf{x}\rangle' \simeq 0$. Hence, if we include a convergence factor $\mathrm{e}^{-\epsilon|\tau|/\hbar}$ in the integrand, for sufficiently small $\epsilon > 0$ we make a negligible difference to the action of $\Omega_\pm$ on a real scattering state: for finite τ this factor approximates unity to arbitrary accuracy as $\epsilon \to 0$. Consider therefore the operator $\widetilde{\Omega}$ that has this harmless factor. Taking the limit that the constant ϵ approaches zero from above we obtain

$$\Omega_\pm|\phi\rangle \longrightarrow \widetilde{\Omega}_\pm|\phi\rangle = \left(1 + \lim_{\epsilon\to 0^+}\frac{\mathrm{i}}{\hbar}\int_0^{\mp\infty}\mathrm{d}\tau\, U^\dagger(\tau) V \mathrm{e}^{-\epsilon|\tau|/\hbar} U_{\mathrm{K}}(\tau)\right)|\phi\rangle. \tag{12.23}$$

The action of $\widetilde{\Omega}_\pm$ on a state $|\phi\rangle$ that is a non-trivial superposition of energy eigenstates is identical to that of $\Omega_\pm$. However, Problem 12.1 shows that the product $H\widetilde{\Omega}_\pm$ satisfies an equation that differs crucially from equation (12.21):

$$H\widetilde{\Omega}_\pm = \widetilde{\Omega}_\pm(H_{\mathrm{K}} \pm \mathrm{i}\epsilon) \mp \mathrm{i}\epsilon. \tag{12.24}$$

Consequently, when we apply H to $|\psi\rangle \equiv \widetilde{\Omega}_\pm|E;\mathrm{free}\rangle$, where $|E;\mathrm{free}\rangle$ is an eigenstate of H_{K}, we find

$$H|\psi\rangle = H\widetilde{\Omega}_\pm|E;\mathrm{free}\rangle = (E \pm \mathrm{i}\epsilon)|\psi\rangle \mp \mathrm{i}\epsilon|E;\mathrm{free}\rangle, \tag{12.25}$$

so $|\psi\rangle$ is an eigenstate of H only when $V = 0$ and $|\psi\rangle = |E;\mathrm{free}\rangle$. Therefore, when we use $\widetilde{\Omega}_\pm$ to generate 'interacting' states from eigenstates of H_{K}, which will henceforth be simply labelled $|E\rangle$, our interacting states are *not* stationary states of the true Hamiltonian, and thus can describe scattering. The crucial point that makes the whole procedure consistent is that for any physically realistic superposition, it makes no difference whether we construct interacting states with $\Omega_\pm$ or $\widetilde{\Omega}_\pm$.

We can simplify $\widetilde{\Omega}_\pm$ a little: since $|\tau| = \tau$ for $\tau \geq 0$ and $|\tau| = -\tau$ for $\tau < 0$, with $|\phi\rangle = |E\rangle$ equation (12.23) becomes

$$\widetilde{\Omega}_\pm|E\rangle = \left(1 + \lim_{\epsilon\to 0^+}\frac{\mathrm{i}}{\hbar}\int_0^{\mp\infty}\mathrm{d}\tau\, U^\dagger(\tau) V \mathrm{e}^{-\mathrm{i}(E\pm\mathrm{i}\epsilon)\tau/\hbar}\right)|E\rangle. \tag{12.26}$$

Therefore, our modification merely supplements the energy eigenvalue E with a small imaginary piece $+\mathrm{i}\epsilon$ for initial states and $-\mathrm{i}\epsilon$ for final states – the sign on $\mathrm{i}\epsilon$ corresponds to the subscript on $\Omega_\pm$ and is historically the origin of the naming of the Ω operators. This procedure is known as the **iϵ prescription**. In practice the prescription is implemented by using the original $\Omega_\pm$ operators, but pretending that all eigenstates of H_{K} satisfy

$$\begin{aligned} H_{\mathrm{K}}|E\rangle &= (E + \mathrm{i}\epsilon)|E\rangle \quad \text{for initial kets when acted on by } \Omega_+ \\ H_{\mathrm{K}}|E'\rangle &= (E' - \mathrm{i}\epsilon)|E'\rangle \quad \text{for final kets when acted on by } \Omega_-. \end{aligned} \tag{12.27a}$$

Similarly, the Hermitian adjoints of the modified operators (12.23) imply that we should likewise pretend that

$$\begin{aligned}\langle E|H_{\rm K} &= \langle E|(E-{\rm i}\epsilon) \quad \text{for initial bras when acted on by } \Omega_+^\dagger \\ \langle E'|H_{\rm K} &= \langle E'|(E'+{\rm i}\epsilon) \quad \text{for final bras when acted on by } \Omega_-^\dagger.\end{aligned} \tag{12.27b}$$

In no way do we mean that the Hermitian operator $H_{\rm K}$ actually has a complex eigenvalues E; equations (12.27) are merely useful fictions that enable us to carry out the iϵ prescription.

12.2.2 Expanding the S-matrix

Since the incoming and outgoing states are free states, and the momentum operators commute with the free Hamiltonian $H_{\rm K}$, the scattering operator is conveniently studied in the momentum representation. We then work with the **S-matrix**

$$\mathcal{S}(\mathbf{p},\mathbf{p}') \equiv \langle\mathbf{p}'|\mathcal{S}|\mathbf{p}\rangle. \tag{12.28}$$

where we must use the iϵ prescription of equations (12.27) to interpret the action of Ω_+ on $|\mathbf{p}\rangle$ that is implicit in this definition. From equation (12.16), the lowest-order contribution to the S-matrix is then

$$\begin{aligned}\langle\mathbf{p}'|\mathcal{S}|\mathbf{p}\rangle &\simeq \langle\mathbf{p}'|\mathbf{p}\rangle - \frac{\rm i}{\hbar}\int_{-\infty}^{\infty}{\rm d}\tau\,\langle\mathbf{p}'|{\rm e}^{{\rm i}H_{\rm K}\tau/\hbar}V{\rm e}^{-{\rm i}H_{\rm K}\tau/\hbar}|\mathbf{p}\rangle \\ &= \langle\mathbf{p}'|\mathbf{p}\rangle - \frac{\rm i}{\hbar}\langle\mathbf{p}'|V|\mathbf{p}\rangle\int_{-\infty}^{\infty}{\rm d}\tau\,{\rm e}^{-{\rm i}(E_p-E_{p'})\tau/\hbar}.\end{aligned} \tag{12.29}$$

Here, we used the rules (12.27) to find that the argument of the exponential in the integrand is actually independent of ϵ. We recognise the integral as $2\pi\hbar\delta(E_p-E_{p'})$.

Potentials that depend only on position are diagonal in the $\mathbf{x}$ representation, so the momentum-space elements $\langle\mathbf{p}'|V|\mathbf{p}\rangle$ are

$$\begin{aligned}\langle\mathbf{p}'|V|\mathbf{p}\rangle &= \int{\rm d}^3\mathbf{x}\,{\rm d}^3\mathbf{x}'\,\langle\mathbf{p}'|\mathbf{x}'\rangle\langle\mathbf{x}'|V|\mathbf{x}\rangle\langle\mathbf{x}|\mathbf{p}\rangle \\ &= \frac{1}{(2\pi\hbar)^3}\int{\rm d}^3\mathbf{x}\,{\rm e}^{-{\rm i}\mathbf{q}\cdot\mathbf{x}/\hbar}V(\mathbf{x})\end{aligned}, \tag{12.30}$$

where we have used our expression (2.78) for the wavefunction of a state of well-defined momentum and defined the **momentum transfer**

$$\mathbf{q} \equiv \mathbf{p}'-\mathbf{p}. \tag{12.31}$$

Therefore, the Born approximation to the S-matrix just depends on the Fourier transform of $V(\mathbf{x})$:

$$\langle\mathbf{p}'|\mathcal{S}^{(1)}|\mathbf{p}\rangle = -\frac{2\pi{\rm i}}{(2\pi\hbar)^3}\delta(E_p-E_{p'})\int{\rm d}^3\mathbf{x}\,{\rm e}^{-{\rm i}\mathbf{q}\cdot\mathbf{x}/\hbar}V(\mathbf{x}). \tag{12.32}$$

From the theory of Fourier transforms, we see that potentials which vary rapidly with $\mathbf{x}$ lead to S-matrices that contain significant amplitudes for large momentum transfers. Turning this around, if a particle suffers a large change in momentum when it is scattered by $V(\mathbf{x})$, we infer that $V(\mathbf{x})$ has sharp features. Arguing along these lines (albeit more classically), Rutherford was able to deduce the existence of nuclei from the occasional back-scattering of α particles off gold foil. More recently, a team of physicists[3] working at SLAC

---

[3] D.H. Coward, *et. al.*, Phys. Rev. Lett. **20**, 292, (1968).

in Stanford scattered high-energy electrons off protons; the electrons sometimes suffered large-angle scattering, providing evidence for the existence of quarks inside the nucleons.

The second-order term in the scattering operator can be treated in a similar manner. From equation (12.16) we find

$$\langle \mathbf{p}'|\mathcal{S}^{(2)}|\mathbf{p}\rangle = -\frac{1}{\hbar^2}\int_{-\infty}^{\infty}\mathrm{d}\tau\int_{-\infty}^{\tau}\mathrm{d}\tau'\,\langle \mathbf{p}'|U_{\mathrm{K}}^{\dagger}(\tau)VU_{\mathrm{K}}(\tau-\tau')VU_{\mathrm{K}}(\tau')|\mathbf{p}\rangle. \tag{12.33}$$

The free-evolution operators can be evaluated by inserting the identity operator $1=\int\mathrm{d}^3\mathbf{k}\,|\mathbf{k}\rangle\langle\mathbf{k}|$ anywhere between the two V operators. Bearing in mind that $H_{\mathrm{K}}|\mathbf{p}\rangle=(E_p+\mathrm{i}\epsilon)|\mathbf{p}\rangle$ and $\langle\mathbf{p}'|H_{\mathrm{K}}=\langle\mathbf{p}'|(E_{p'}+\mathrm{i}\epsilon)$ in accordance with the iϵ prescription, we find

$$\begin{aligned}\langle \mathbf{p}'|\mathcal{S}^{(2)}|\mathbf{p}\rangle = \lim_{\epsilon\to0^+} -\frac{1}{\hbar^2}\int\mathrm{d}^3\mathbf{k}\Big(&\langle\mathbf{p}'|V|\mathbf{k}\rangle\langle\mathbf{k}|V|\mathbf{p}\rangle\\ &\times\int_{-\infty}^{\infty}\mathrm{d}\tau\int_{-\infty}^{\tau}\mathrm{d}\tau'\,\mathrm{e}^{\mathrm{i}(E_{p'}+\mathrm{i}\epsilon-E_k)\tau/\hbar}\mathrm{e}^{-\mathrm{i}(E_p-E_k+\mathrm{i}\epsilon)\tau'/\hbar}\Big).\end{aligned} \tag{12.34}$$

The integral over τ' is

$$\int_{-\infty}^{\tau}\mathrm{d}\tau'\;\mathrm{e}^{-\mathrm{i}(E_p-E_k+\mathrm{i}\epsilon)\tau'/\hbar}=\mathrm{i}\hbar\frac{\mathrm{e}^{-\mathrm{i}(E_p-E_k+\mathrm{i}\epsilon)\tau/\hbar}}{E_p-E_k+\mathrm{i}\epsilon}, \tag{12.35}$$

so the second-order contribution to the S-matrix is

$$\begin{aligned}\mathcal{S}^{(2)}(\mathbf{p}',\mathbf{p}) &= \lim_{\epsilon\to0^+}-\frac{\mathrm{i}}{\hbar}\int\mathrm{d}^3\mathbf{k}\,\frac{\langle\mathbf{p}'|V|\mathbf{k}\rangle\langle\mathbf{k}|V|\mathbf{p}\rangle}{E_p-E_k+\mathrm{i}\epsilon}\int_{-\infty}^{\infty}\mathrm{d}\tau\,\mathrm{e}^{-\mathrm{i}(E_p-E_{p'})\tau/\hbar}\\ &=-2\pi\mathrm{i}\,\delta(E_p-E_{p'})\lim_{\epsilon\to0^+}\int\mathrm{d}^3\mathbf{k}\,\frac{\langle\mathbf{p}'|V|\mathbf{k}\rangle\langle\mathbf{k}|V|\mathbf{p}\rangle}{E_p-E_k+\mathrm{i}\epsilon}.\end{aligned} \tag{12.36}$$

The numerator in the integrand is the amplitude for the particle to scatter from the $|\mathbf{p}\rangle$ state into the $|\mathbf{k}\rangle$ state, and then from the $|\mathbf{k}\rangle$ state into the $|\mathbf{p}'\rangle$ state. The denominator arose from the integration over τ', which in turn was present because the particle travelled freely for some time $\tau-\tau'$ in between the two interactions. Since it comes from this free propagation, the factor $(E_p-E_k+\mathrm{i}\epsilon)^{-1}$ is known as the **propagator**, written here in the momentum representation. Finally, because we do not measure the intermediate state, equation (12.36) adds up the amplitudes for scattering via any state.

Higher order terms are handled in a similar way: V occurs n times in $\mathcal{S}^{(n)}(\mathbf{p}',\mathbf{p})$, so there are $n-1$ intermediate evolution operators, leading to $n-1$ propagators. Similarly, there are $n-1$ sets of intermediate states, all of which are integrated over. $\mathcal{S}^{(n)}(\mathbf{p}',\mathbf{p})$ may be represented diagrammatically as in Figure 12.1. These **Feynman diagrams** are an order-by-order bookkeeping system for calculating contributions to the S-matrix: each term in the series for the S-matrix corresponds to a diagram, and **Feynman rules** can be defined that enable the algebraic expression for the term to be inferred from the diagram. Thus Feynman diagrams summarise complicated integrals in an intuitive way.

The Feynman rules required here are extremely simple: (i) each vertex has just two lines going into it and is associated with a factor V; (ii) each 'internal line' (one that has a vertex at each end) is associated with the propagator $(E_p-E_k+\mathrm{i}\epsilon)^{-1}$, where $\mathbf{k}$, which is integrated over, is the momentum carried by that line; (iii) there is an overall prefactor $-2\pi\mathrm{i}\delta(E_p-E_{p'})$, where $\mathbf{p}$ and $\mathbf{p}'$ are the ingoing and outgoing momenta, respectively. With these rules we can only construct one diagram with a given number n of vertices, and it's a simple chain. Feynman diagrams become much more interesting and valuable when one recognises that when an electron is scattered by an

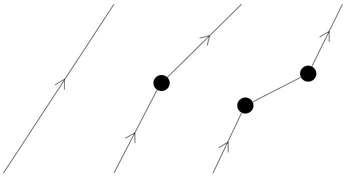


Figure 12.1 Feynman diagrams for the scattering process to lowest orders in V.

electrostatic potential $V(\mathbf{x})$, for example, it really collides with a photon, and one needs to include the coupled dynamics of the photons. In this more sophisticated picture, $V(\mathbf{x})$ is replaced by the electromagnetic vector potential $\mathbf{A}$, which becomes a quantum-mechanical object, and our diagrams include propagators for both photons and electrons. Moreover, the vertices become points at which three or more lines meet, two for the incoming and outgoing electron, and one or more for photons. With a richer set of lines and vertices on hand, many different diagrams can be constructed that all have the same number of vertices, and therefore contribute to the S-matrix at the same order.

12.2.3 The scattering amplitude

Both the first- and second-order approximations to the S-matrix are proportional to an energy-conserving delta function. This result is not limited to the series expansion for $\mathcal{S}$, but actually holds for the exact S-matrix as we now demonstrate. Equation (12.20) and its Hermitian adjoint state that $H\Omega_\pm = \Omega_\pm H_\mathrm{K}$ and $\Omega_\pm^\dagger H = H_\mathrm{K}\,\Omega_\pm^\dagger$. Now $\mathcal{S} \equiv \Omega_-^\dagger \Omega_+$, so

$$H_\mathrm{K}\mathcal{S} = H_\mathrm{K}\,\Omega_-^\dagger\Omega_+ = \Omega_-^\dagger H\Omega_+ = \Omega_-^\dagger\Omega_+ H_\mathrm{K} = \mathcal{S}H_\mathrm{K}, \tag{12.37}$$

that is, $[\mathcal{S}, H_\mathrm{K}] = 0$. Sandwiching this commutation relation between momentum eigenstates and using the iϵ prescription of equations (12.27) gives the relation

$$0 = \langle\mathbf{p}'|[\mathcal{S}, H_\mathrm{K}]|\mathbf{p}\rangle = (E_p + \mathrm{i}\epsilon - E_{p'} - \mathrm{i}\epsilon)\mathcal{S}(\mathbf{p}, \mathbf{p}') = (E_p - E_{p'})\mathcal{S}(\mathbf{p}, \mathbf{p}'), \tag{12.38}$$

so the S-matrix vanishes unless the initial and final states have the same (real) energy. This tells us that the exact S-matrix must have the form $\mathcal{S}(\mathbf{p}, \mathbf{p}') \propto \delta(E_p - E_{p'})$.

In equation (12.6) we broke $\mathcal{S}$ into the sum $\mathcal{S} = 1 + \mathcal{T}$ to isolate the scattering amplitude, and it is clear that $\langle\mathbf{p}'|\mathcal{T}|\mathbf{p}\rangle$ is also proportional to $\delta(E_p - E_{p'})$. Motivated by this insight we define the **scattering amplitude** $f(\mathbf{p} \to \mathbf{p}')$ by

$$\langle\mathbf{p}'|\mathcal{T}|\mathbf{p}\rangle = \frac{\mathrm{i}}{2\pi\hbar m}\delta(E_p - E_{p'})f(\mathbf{p} \to \mathbf{p}'), \tag{12.39}$$

where the factor of $\mathrm{i}/(2\pi\hbar m)$ is included for later convenience. On account of the delta function, $f(\mathbf{p} \to \mathbf{p}')$ depends on $\mathbf{p}'$ only through its direction $\hat{\mathbf{p}}'$.

To understand the significance of the scattering amplitude, consider the following argument. According to the discussion in §12.1, long after the interaction, a particle that scattered from the free state $|\phi\rangle$ can be described by the free state $|\lambda\rangle = \mathcal{S}|\phi\rangle$. Therefore, in the idealised case that the initial state was a momentum eigenstate $|\mathbf{p}\rangle$, the wavefunction of the final state is

$$\begin{aligned}\langle\mathbf{r}|\lambda\rangle &= \langle\mathbf{r}|\mathcal{S}|\mathbf{p}\rangle = \langle\mathbf{r}|\mathbf{p}\rangle + \int \mathrm{d}^3\mathbf{p}'\,\langle\mathbf{r}|\mathbf{p}'\rangle\langle\mathbf{p}'|\mathcal{T}|\mathbf{p}\rangle \\ &= \langle\mathbf{r}|\mathbf{p}\rangle + \frac{\mathrm{i}}{2\pi\hbar m}\int \mathrm{d}^3\mathbf{p}'\,\langle\mathbf{r}|\mathbf{p}'\rangle\delta(E_p - E_{p'})f(\mathbf{p} \to \mathbf{p}').\end{aligned} \tag{12.40}$$

Since the states $|\mathbf{p}'\rangle$ in the integrand are final states, the $\mathrm{i}\epsilon$ prescription tells us to take ${p'}^2/2m = (E_{p'} - \mathrm{i}\epsilon)$, so in spherical polar coordinates[4]

$$\mathrm{d}^3\mathbf{p}' = {p'}^2\mathrm{d}p'\mathrm{d}\Omega = m\sqrt{2m(E_{p'} - \mathrm{i}\epsilon)}\,\mathrm{d}E_{p'}\mathrm{d}\Omega. \tag{12.41}$$

Using this in equation (12.40) and integrating over $E_{p'}$ using the delta function gives

$$\langle\mathbf{r}|\lambda\rangle = \langle\mathbf{r}|\mathbf{p}\rangle + \frac{\mathrm{i}\sqrt{2m(E_p - \mathrm{i}\epsilon)}}{(2\pi\hbar)^{5/2}}\int\mathrm{d}\Omega\,\mathrm{e}^{\mathrm{i}r\sqrt{2m(E_p-\mathrm{i}\epsilon)}\;\hat{\mathbf{p}}'\cdot\hat{\mathbf{r}}/\hbar}f(\mathbf{p}\to\mathbf{p}'). \tag{12.42}$$

This free-particle wavefunction only looks like the true wavefunction of the scattered particle long after the collision, so equation (12.42) will only correspond to the physical wavefunction as $r\to\infty$. In this limit, the phase of the exponential in equation (12.42) varies extremely rapidly as a function of the variables θ and ϕ that define the direction of $\hat{\mathbf{p}}$, over which we are integrating. The different contributions to the integral will therefore cancel each other out except where the phase of the integrand is stationary with respect to angle. For sufficiently large r the sensitivity of the exponential to angle will exceed that of $f(\mathbf{p}\to\mathbf{p}')$. Hence the dominant contribution to the integral arises when

$$\frac{\partial}{\partial\theta}(\hat{\mathbf{p}}'\cdot\hat{\mathbf{r}}) = \frac{\partial}{\partial\theta}\cos\theta = 0 \quad\text{and}\quad \frac{\partial}{\partial\phi}\cos\theta = 0, \tag{12.43}$$

where we have aligned the polar axis with the (fixed) direction $\hat{\mathbf{r}}$. These conditions are satisfied when $\theta = 0, \pi$, independent of ϕ. When $\theta = \pi$, and $\hat{\mathbf{p}}'\cdot\hat{\mathbf{r}} = -1$, the integrand of equation (12.42) is exponentially suppressed as $r\to\infty$ by the $\mathrm{i}\epsilon$ prescription. Therefore the integral over the unit sphere is dominated by the contribution from a small disc centred on the direction $\hat{\mathbf{r}}$. This insight justifies the approximation

$$\begin{aligned}\int&\mathrm{d}\Omega\,\mathrm{e}^{\mathrm{i}r\sqrt{2m(E_p-\mathrm{i}\epsilon)}\;\hat{\mathbf{p}}'\cdot\hat{\mathbf{r}}/\hbar}f(\mathbf{p}\to\mathbf{p}')\\ &\simeq 2\pi f(\mathbf{p}\to\mathbf{p}')\int\mathrm{d}\cos\theta\,\mathrm{e}^{\mathrm{i}r\sqrt{2m(E_p-\mathrm{i}\epsilon)}\cos\theta/\hbar}\\ &\simeq 2\pi\hbar\frac{f(\mathbf{p}\to\mathbf{p}')}{\mathrm{i}r\sqrt{2m(E_p-\mathrm{i}\epsilon)}}\mathrm{e}^{\mathrm{i}r\sqrt{2m(E_p-\mathrm{i}\epsilon)}\cos\theta/\hbar}.\end{aligned} \tag{12.44}$$

Using this expression in equation (12.42) we have finally

$$\begin{aligned}\lim_{r\to\infty}\langle\mathbf{r}|\phi\rangle &= \lim_{r\to\infty}\left(\langle\mathbf{r}|\mathbf{p}\rangle + \frac{1}{(2\pi\hbar)^{3/2}}\frac{\mathrm{e}^{\mathrm{i}r\sqrt{2m(E_p-\mathrm{i}\epsilon)}/\hbar}}{r}f(\mathbf{p}\to p\hat{\mathbf{r}})\right)\\ &= \lim_{r\to\infty}\frac{1}{(2\pi\hbar)^{3/2}}\left(\mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{r}/\hbar} + \frac{\mathrm{e}^{\mathrm{i}pr/\hbar}}{r}f(\mathbf{p}\to p\hat{\mathbf{r}})\right),\end{aligned} \tag{12.45}$$

where in the last line we have taken the limit $\epsilon\to0^+$. Equation (12.45) shows that a particle that was initially in a momentum eigenstate will emerge from the scattering process in a superposition of its original state (no scattering) and a wave travelling radially outwards. The scattering amplitude $f(\mathbf{p}\to p\hat{\mathbf{r}})$ is just the amplitude of this outgoing wave.

In equation (12.45) the time-dependence is suppressed by our convention that the S-matrix generates the wavefunction at the generic time 0. We now restore explicit time dependence by introducing a factor $\mathrm{e}^{-\mathrm{i}E_pt/\hbar}$ and replace the incoming state $|\mathbf{p}\rangle$ by a realistic superposition of such states. Then the outgoing wavefunction becomes

$$\langle\mathbf{r}|\phi;t\rangle = \int\frac{\mathrm{d}^3\mathbf{p}\,\phi(\mathbf{p})}{(2\pi\hbar)^{3/2}}\left(\mathrm{e}^{\mathrm{i}(\mathbf{p}\cdot\mathbf{r}-E_pt)/\hbar} + \frac{\mathrm{e}^{\mathrm{i}(pr-E_pt)/\hbar}}{r}f(\mathbf{p}\to p\hat{\mathbf{r}})\right), \tag{12.46}$$

which is the sum of the incoming wave packet plus a wave packet that travels radially outwards from the scattering centre.

---

[4] In this chapter it is convenient to define $\mathrm{d}\Omega = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$ rather than $\mathrm{d}^2\Omega = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$ as in earlier chapters.

12.3 Cross-sections and scattering experiments

Children sometimes test their skill by taking turns to throw pebbles at a distant target, perhaps a rock. If a pebble hits, it will bounce off in a different direction, whereas a pebble that misses will simply continue undisturbed. Each throw will not be repeated exactly, and after a long time we might imagine that the children have thrown pebbles randomly, such that the distribution of throws per unit area is uniform over a region surrounding the target. If so, we can estimate the area of the target that the children see by simply counting the number of pebbles that hit it – if $N_{\rm in}$ pebbles are thrown in per unit area, and $N_{\rm sc}$ of them hit the rock, the rock has cross-sectional area

$$A \simeq N_{\rm sc}/N_{\rm in}. \tag{12.47}$$

With more care, we can measure the angle through which throws are deflected. Pebbles that strike nearby points of a smooth rock will bounce off in roughly the same direction, whereas a jagged rock may deflect pebbles that hit closely spaced points very differently. Hence, counting the number of pebbles that end up going in a given direction gives us information about the rock's shape. We define the **differential cross-section** $\delta\sigma$ to be the area of the target that deflects pebbles into a small solid angle $\delta\Omega$. If there are $N(\theta,\phi)\delta\Omega$ such pebbles, then

$$\delta\sigma \equiv \frac{N(\theta,\phi)\delta\Omega}{N_{\rm in}} \qquad \text{or} \qquad \frac{\delta\sigma}{\delta\Omega} = \frac{N(\theta,\phi)}{N_{\rm in}}, \tag{12.48}$$

and the total cross-section is

$$\sigma_{\rm tot} \equiv \int {\rm d}\sigma = \int {\rm d}\Omega\,\frac{{\rm d}\sigma}{{\rm d}\Omega} = \int {\rm d}\Omega\,\frac{N(\theta,\phi)}{N_{\rm in}} = \frac{N_{\rm sc}}{N_{\rm in}} \tag{12.49}$$

as above.

This may seem a rather baroque manner in which to investigate rocks, but when you go out on a dark night with a torch, you probe objects in a very similar way by throwing photons at them. A more complete analogy can be drawn between pebble-throwing children and physicists with particle accelerators: a beam containing a large number $N_{\rm b}$ of particles is fired towards a target, and detectors measure the number of particles that scatter off into each element of solid angle $\delta\Omega$. Long before the collision, a typical particle in the beam looks like a free state $|\phi\rangle$, so the probability density of each particle is $|\langle\mathbf{x}|\phi\rangle|^2$ and the number of particles per unit area perpendicular to the beam direction is

$$n_{\rm in}(\mathbf{x}_\perp) = N_{\rm b}\int {\rm d}x_\parallel\, |\langle\mathbf{x}|\phi\rangle|^2, \tag{12.50}$$

where the integral is along the beam direction.

When $|\phi\rangle$ is expanded in terms of momentum eigenstates, equation (12.50) becomes

$$\begin{aligned} n_{\rm in}(\mathbf{x}_\perp) &= \frac{N_{\rm b}}{(2\pi\hbar)^3}\int {\rm d}x_\parallel {\rm d}^3\mathbf{p}\,{\rm d}^3\mathbf{p}'{\rm e}^{{\rm i}(\mathbf{p}-\mathbf{p}')\cdot\mathbf{x}/\hbar}\phi(\mathbf{p})\phi^*(\mathbf{p}') \\ &= \frac{N_{\rm b}}{(2\pi\hbar)^2}\int {\rm d}^3\mathbf{p}\,{\rm d}^3\mathbf{p}'\,\delta(p_\parallel - p'_\parallel){\rm e}^{{\rm i}(\mathbf{p}_\perp-\mathbf{p}'{}_\perp)\cdot\mathbf{x}_\perp/\hbar}\phi(\mathbf{p})\phi^*(\mathbf{p}'), \end{aligned} \tag{12.51}$$

where the integral over $x_\parallel$ produced the delta function of momentum along the beam direction. Experimental beams are highly collimated, so $\phi(\mathbf{p})$ vanishes rapidly unless the momentum is near some average value $\bar{\mathbf{p}}$. In particular, they contain only small amounts of momentum perpendicular to the beam direction $\hat{\bar{\mathbf{p}}} \equiv \bar{\mathbf{p}}/|\bar{\mathbf{p}}|$. Consequently, throughout a region of non-negligible extent near the centre of the beam, at $\mathbf{x}_\perp = 0$, we have ${\rm e}^{{\rm i}\mathbf{p}_\perp\cdot\mathbf{x}_\perp/\hbar} \simeq$

1. With this approximation, the number of particles incident per unit area perpendicular to the beam is uniform near the beam centre, so

$$n_{\rm in}(\mathbf{x}_\perp) \simeq \frac{N_{\rm b}}{(2\pi\hbar)^2}\int \mathrm{d}^3\mathbf{p}\,\mathrm{d}^3\mathbf{p}'\,\delta(p_\parallel - p'_\parallel)\phi(\mathbf{p})\phi^*(\mathbf{p}'). \tag{12.52}$$

Equation (12.52) may seem a bizarre way to rewrite the intuitively clear expression (12.50), but it will soon prove its worth.

We must now calculate $N(\theta,\phi)$. At large distances, we know that the wavefunction of particles scattered from the state $|\phi\rangle$ is $\langle\mathbf{r}|\mathcal{T}|\phi\rangle$. If we had placed detectors at some large distance r_0 from the scattering centre, over time they would have detected any particle that has the same values of θ, ϕ and is predicted to lie at $r > r_0$. Thus the total number of particles that are detected in the element of solid angle $\delta\Omega$ is

$$N(\theta,\phi)\delta\Omega = \int_{r_0}^{\infty}\mathrm{d}r\, r^2 N_{\rm b}\delta\Omega\,|\langle\mathbf{r}|\mathcal{T}|\phi\rangle|^2. \tag{12.53}$$

Equation (12.46) gives $\langle\mathbf{r}|\phi\rangle = \langle\mathbf{r}|(1+\mathcal{T})|\phi\rangle$, so

$$N(\theta,\phi)\delta\Omega = N_{\rm b}\delta\Omega\int_{r_0}^{\infty}\mathrm{d}r\,\left|\int\frac{\mathrm{d}^3\mathbf{p}\,\phi(\mathbf{p})}{(2\pi\hbar)^{3/2}}\,\mathrm{e}^{\mathrm{i}pr/\hbar}f(\mathbf{p}\to p\hat{\mathbf{r}})\right|^2. \tag{12.54}$$

So long as the scattering amplitude is reasonably smooth, the collimation of the beam allows us to replace $f(\mathbf{p}\to p\hat{\mathbf{r}})$ by its value at the average momentum $f(\bar{\mathbf{p}}\to\bar{p}\hat{\mathbf{r}})$, which gives

$$N(\theta,\phi)\delta\Omega = N_{\rm b}\delta\Omega|f(\bar{\mathbf{p}}\to\bar{p}\hat{\mathbf{r}})|^2\int\frac{\mathrm{d}^3\mathbf{p}\,\mathrm{d}^3\mathbf{p}'}{(2\pi\hbar)^3}\phi(\mathbf{p})\phi^*(\mathbf{p}')\int_{r_0}^{\infty}\mathrm{d}r\,\mathrm{e}^{\mathrm{i}(p-p')r/\hbar}. \tag{12.55}$$

Explicitly writing the incoming particle's momentum in terms of the average momentum $\bar{\mathbf{p}}$ of the beam and its deviation $\delta\mathbf{p}$ from this value, we find

$$p = \sqrt{(\bar{\mathbf{p}}+\delta\mathbf{p})\cdot(\bar{\mathbf{p}}+\delta\mathbf{p})} \simeq \bar{p}\left(1+\frac{\bar{\mathbf{p}}\cdot\delta\mathbf{p}}{\bar{p}^2}\right) = \bar{p}+\hat{\bar{\mathbf{p}}}\cdot\delta\mathbf{p}, \tag{12.56}$$

so the argument of the exponential in equation (12.55) involves

$$(p-p') \simeq (\delta\mathbf{p}-\delta\mathbf{p}')\cdot\hat{\bar{\mathbf{p}}} = (\delta p_\parallel - \delta p'_\parallel) = (p_\parallel - p'_\parallel). \tag{12.57}$$

Since $r > r_0$ is very large, the phase of this exponential oscillates rapidly, so again the integral is dominated by contributions for which $p_\parallel = p'_\parallel$ giving

$$\begin{aligned} N(\theta,\phi)\delta\Omega &\simeq N_{\rm b}\,\delta\Omega|f(\bar{\mathbf{p}}\to\bar{p}\hat{\mathbf{r}})|^2\int\frac{\mathrm{d}^3\mathbf{p}\mathrm{d}^3\mathbf{p}'}{(2\pi\hbar)^2}\,\phi(\mathbf{p})\phi^*(\mathbf{p}')\delta(p_\parallel - p'_\parallel) \\ &= n_{\rm in}|f(\bar{\mathbf{p}}\to\bar{p}\hat{\mathbf{r}})|^2\delta\Omega, \end{aligned} \tag{12.58}$$

where we have used equation (12.52).

Combining this with the definition (12.48) of the differential cross-section, we find $\mathrm{d}\sigma/\mathrm{d}\Omega$ for scattering from momentum $\mathbf{p}$ (now relabelled) into a different momentum $\mathbf{p}'$ of the same magnitude:[5]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = |f(\mathbf{p}\to\mathbf{p}')|^2, \tag{12.59}$$

---

[5] Our language here is loose: neither the incoming nor the outgoing states are strictly states of well-defined momentum.

where $\hat{\mathbf{p}}'$ points towards the centre of the element of solid angle $\mathrm{d}\Omega$. The total scattering cross-section is

$$\sigma_{\rm tot} = \int \mathrm{d}\Omega\, |f(\mathbf{p}\to\mathbf{p}')|^2. \tag{12.60}$$

The two remarkably simple formulae (12.59) and (12.60) form crucial links between experiment and theory. If the scattering potential is sufficiently weak that the Born approximation is valid, equation (12.32) tells us that the scattering amplitude is $f(\mathbf{p}\to\mathbf{p}') = -4\pi^2\hbar m\langle\mathbf{p}'|V|\mathbf{p}\rangle$, and the differential cross-section is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = (4\pi^2\hbar m)^2|\langle\mathbf{p}'|V|\mathbf{p}\rangle|^2 = \left|\frac{m}{2\pi\hbar^2}\int \mathrm{d}^3\mathbf{x}\,\mathrm{e}^{-\mathrm{i}\mathbf{q}\cdot\mathbf{x}/\hbar}V(\mathbf{x})\right|^2, \tag{12.61}$$

where $\mathbf{q} = \mathbf{p}' - \mathbf{p}$. The integral in equation (12.61) is just the Fourier transform $\widetilde{V}(\mathbf{q})$ of the potential, so the equation can be rewritten

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{m^2}{4\pi^2\hbar^4}P(q), \tag{12.62}$$

where $P(\mathbf{q}) = |\widetilde{V}(\mathbf{q})|^2$ is the **power spectrum** of $V(\mathbf{x})$. Thus, by measuring the number of particles that are scattered into a given direction, we can determine the power spectrum of the interaction potential.

If we could complement this information by measuring the phases of the Fourier transform, we could reconstruct $V(\mathbf{x})$ from the scattering data. The obvious way to measure the phases is to observe interference between the scattered and incident amplitudes – interference of this type is what generates holograms, from which the three-dimensional structure of the scattering object can be reconstructed. A high-energy accelerator does not produce sufficiently pure quantum states for interference between the incident and scattered amplitudes to be observable. Moreover, in realistic circumstances, experiments in which this interference was observed would be of limited interest because in reality the potential $V(\mathbf{x})$ fluctuates in time. For example, in §12.4 below we discuss scattering of electrons by atoms, and in this case the electrostatic potential V varies in time as the electrons that partly generate it whizz about the atom. These internal motions cause rapid variability in the phases of $\widetilde{V}(\mathbf{q})$, while affecting the power spectrum of V to a much smaller extent: the latter depends on the number and structure of the lumps associated with the electrons and nucleus, rather than on their locations. Thus scattering experiments enable us to unveil as much of the structure of matter as we are in practice interested in. For this reason they are one of the most powerful tools we can deploy in our efforts to understand nature.

12.3.1 The optical theorem

The simple connection between the power spectrum of $V(\mathbf{x})$ and the scattering cross-section established above relies on the Born approximation. This approximation is certainly not always valid, so it is interesting to see what we can say about cross-sections in general.

From equation (12.8) we have that

$$\mathcal{T} + \mathcal{T}^\dagger = -\mathcal{T}^\dagger\mathcal{T} = -\int \mathrm{d}^3\mathbf{p}''\,\mathcal{T}^\dagger|\mathbf{p}''\rangle\langle\mathbf{p}''|\mathcal{T}. \tag{12.63}$$

Squeezing this equation between $\langle\mathbf{p}'|$ and $|\mathbf{p}\rangle$ and using equation (12.39), we find that the scattering amplitude $f(\mathbf{p}\to\mathbf{p}')$ satisfies

$$\begin{aligned}&\delta(E_p - E_{p'})\big\{f(\mathbf{p}\to\mathbf{p}') - f^*(\mathbf{p}'\to\mathbf{p})\big\}\\ &\quad = \frac{\mathrm{i}}{2\pi\hbar m}\int \mathrm{d}^3\mathbf{p}''\,\delta(E_{p''} - E_{p'})f^*(\mathbf{p}'\to\mathbf{p}'')\,\delta(E_{p''} - E_p)f(\mathbf{p}\to\mathbf{p}'').\end{aligned} \tag{12.64}$$

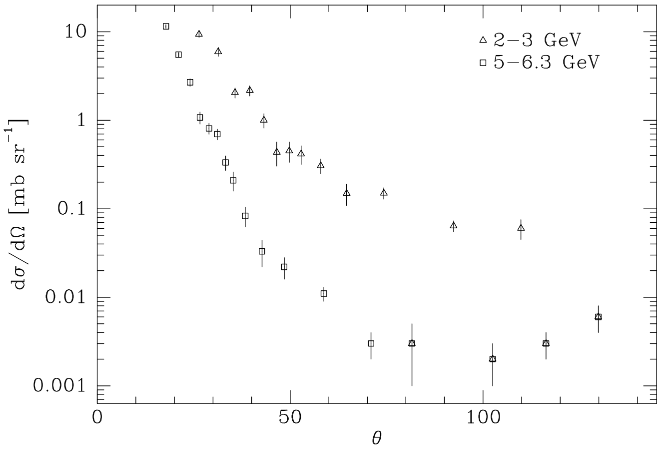


Figure 12.2 The differential cross-section for neutron-proton scattering at two values of the centre-of-mass energy. Data obtained from M. Kreisler *et. al.*, Phys. Rev. Lett., **16**, 1217, (1966). The diffraction peak at $\theta = 0$ can be understood in terms of the optical theorem.

The second delta function in the integral ensures that $E_{p''} = E_p$, so we can replace the first delta function by $\delta(E_p - E_{p'})$, and then bring it outside the integral since it no longer depends on $\mathbf{p}''$. Then we have

$$f(\mathbf{p} \to \mathbf{p}') - f^*(\mathbf{p}' \to \mathbf{p}) = \frac{\mathrm{i}}{2\pi\hbar m} \int \mathrm{d}^3\mathbf{p}''\, \delta(E_{p''} - E_p) f^*(\mathbf{p}' \to \mathbf{p}'') f(\mathbf{p} \to \mathbf{p}''). \tag{12.65}$$

When $\mathbf{p}' = \mathbf{p}$ (equal directions as well as magnitudes), the left side becomes $f(\mathbf{p} \to \mathbf{p}) - f^*(\mathbf{p} \to \mathbf{p}) = 2\mathrm{i}\Im\mathrm{m} f(\mathbf{p} \to \mathbf{p})$ and, after changing variables in the delta function to obtain $\delta(E_{p''} - E_p) = (m/p'')\,\delta(p'' - p)$, equation (12.65) reduces to

$$\begin{aligned} \Im\mathrm{m} f(\mathbf{p} \to \mathbf{p}) &= \frac{1}{4\pi\hbar} \int \mathrm{d}^3\mathbf{p}''\, \frac{1}{p''} \delta(p - p'') f^*(\mathbf{p} \to \mathbf{p}'') f(\mathbf{p} \to \mathbf{p}'') \\ &= \frac{p}{4\pi\hbar} \int \mathrm{d}\Omega\, |f(\mathbf{p} \to \mathbf{p}'')|^2. \end{aligned} \tag{12.66}$$

In this last expression we recognise from equation (12.60) the total cross-section for scattering from a state of initial momentum $\mathbf{p}$. We have derived the relation

$$\sigma_{\mathrm{tot}}(\mathbf{p}) = \frac{4\pi\hbar}{p} \Im\mathrm{m} f(\mathbf{p} \to \mathbf{p}). \tag{12.67}$$

This equation is known as the **optical theorem**, and relates the total cross-section to the imaginary part of the scattering amplitude in the forward direction. It is at heart a re-expression of equation (12.9) with an identity operator $\sum |\psi_i\rangle\langle\psi_i|$ inserted after $\mathcal{T}^\dagger$ on the right. The forward scattering gives the probability a particle is removed from the original beam, and this is associated with the total probability the particle is deflected into some other direction.

When neutrons are scattered from protons, the differential cross-section has a peak in the forward direction, as shown in Figure 12.2. As the centre-of-mass energy is raised, this peak increases in height and decreases in width. This behaviour is explained by the optical theorem. Experimentally, the total cross-section becomes roughly constant as $p \to \infty$. Equation (12.67) then implies that $\Im\mathrm{m} f(\mathbf{p} \to \mathbf{p})$ rises roughly in proportion to p, so from equation (12.59) the differential cross-section in the forward direction grows at least as fast as p^2. Conversely, since $|f(\mathbf{p} \to \mathbf{p}')|^2$ is necessarily positive, the total cross section $\sigma_{\mathrm{tot}} = \int \mathrm{d}\Omega\, |f(\mathbf{p} \to \mathbf{p}')|^2$ is never less than the cross-section for scattering into any solid angle $\Delta\Omega < 4\pi$. Choosing $\Delta\Omega$ to be the region around the forward direction in which $|f(\mathbf{p}\to\mathbf{p}')|^2$ is falling from its peak at $\mathbf{p}'=\mathbf{p}$, but still greater than $\frac{1}{2}|f(\mathbf{p}\to\mathbf{p})|^2$, gives

$$\sigma_{\rm tot}\geq\int_{\Delta\Omega}{\rm d}\Omega\,|f(\mathbf{p}\to\mathbf{p}')|^2\geq\tfrac{1}{2}\Delta\Omega|f(\mathbf{p}\to\mathbf{p})|^2\geq\tfrac{1}{2}\Delta\Omega|\Im{\rm m}f(\mathbf{p}\to\mathbf{p})|^2. \tag{12.68}$$

Hence, from the optical theorem, the FWHM of the peak around the forward direction is bounded by

$$\Delta\Omega\leq\frac{32\pi^2}{p^2\sigma_{\rm tot}} \tag{12.69}$$

and therefore shrinks as p^{-2} as $p\to\infty$. This **diffraction peak** is familiar from optics: collimated light can be diffracted by two slits, and the resulting intensity in the Fraunhofer region is peaked in the forward direction, with a FWHM that shrinks as the frequency of the light is increased.

12.4 Scattering electrons off hydrogen

We now apply our scattering formalism to a physical problem, namely scattering of electrons by a hydrogen atom that is in its ground state $|1,0,0\rangle$ (§8.1). Taking the proton to be a pointlike object at the centre of the atom, the atom's charge distribution is

$$\rho(\mathbf{r})=e\delta^3(\mathbf{r})-e\,|\langle\mathbf{r}|1,0,0\rangle|^2\,. \tag{12.70}$$

From §8.1.2 we have that $|\langle\mathbf{r}|1,0,0\rangle|^2={\rm e}^{-2r/a_0}/\pi a_0^3$ where a_0 is the Bohr radius (eq. 8.13b). Hence, the atom is the source of an electric field $\mathbf{E}=-\nabla\Phi$, where

$$\begin{aligned}\Phi(\mathbf{r})&=\frac{1}{4\pi\epsilon_0}\int{\rm d}^3\mathbf{r}'\,\frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}=\frac{e}{4\pi\epsilon_0 r}-\frac{e}{4\pi\epsilon_0\pi a_0^3}\int{\rm d}^3\mathbf{r}'\,\frac{{\rm e}^{-2r'/a_0}}{|\mathbf{r}-\mathbf{r}'|}\\&=\frac{e}{4\pi\epsilon_0 r}-\frac{2e}{4\pi a_0^3\epsilon_0}\int{\rm d}r'{\rm d}\theta\,\frac{r'^2\sin\theta\,{\rm e}^{-2r'/a_0}}{(r^2+r'^2-2rr'\cos\theta)^{1/2}}.\end{aligned} \tag{12.71}$$

The integral differs only trivially from that evaluated in Box 10.1. Adapting the result obtained there we conclude that

$$\Phi(\mathbf{r})=\frac{e}{4\pi\epsilon_0}\left(\frac{1}{r}+\frac{1}{a_0}\right){\rm e}^{-2r/a_0}. \tag{12.72}$$

Notice how the ground-state electron shields the pure $1/r$ Coulomb potential of the proton, causing the overall potential to decline exponentially at large distances. This potential will scatter a passing charged particle such as an electron. It will turn out that our calculations only apply to electrons that have enough energy to excite or even ionise the atom. Never the less, we shall consider only the case of elastic scattering, in which the atom remains throughout in its ground state.

Equation (12.61) gives the Born approximation for the differential cross section in terms of the Fourier transform of the interaction potential $V(\mathbf{r})=-e\Phi(\mathbf{r})$. By equation (12.72) V is a function of distance r only, and for any such function it is straightforward to show that

$$\int{\rm d}^3\mathbf{r}\,{\rm e}^{-{\rm i}\mathbf{q}\cdot\mathbf{r}/\hbar}V(r)=-\frac{4\pi\hbar}{q}\int_0^\infty{\rm d}r\,r\sin\left(\frac{qr}{\hbar}\right)V(r). \tag{12.73}$$

Substituting for $V(r)=-e\Phi(r)$ from equation (12.72) we find

$$\int{\rm d}^3\mathbf{r}\,{\rm e}^{-{\rm i}\mathbf{q}\cdot\mathbf{r}/\hbar}V(r)=\frac{4\pi\hbar^2}{m_{\rm e}}\frac{8+(qa_0/\hbar)^2}{(4+(qa_0/\hbar)^2)^2}. \tag{12.74}$$

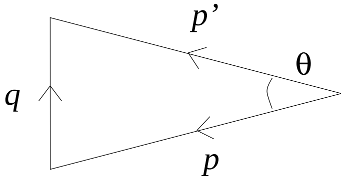


Figure 12.3 Trigonometry of the isosceles triangle tells us the magnitude of momentum transfer, since $|\mathbf{p}'| = |\mathbf{p}|$.

Plugging this result into equation (12.61), we have finally

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = 4a_0^2\left(\frac{8+(qa_0/\hbar)^2}{(4+(qa_0/\hbar)^2)^2}\right)^2. \tag{12.75}$$

Now $q = |\mathbf{p}' - \mathbf{p}| = 2p\sin(\theta/2)$ (see Figure 12.3), so q is smallest and the cross-section is greatest for forward scattering ($\theta = 0$). Quantitatively,

$$\left.\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right|_{\theta=0} = a_0^2, \tag{12.76}$$

independent of the incoming electron's energy. When the electron's momentum is large, the cross-section drops sharply as we move away from the forward direction. This behaviour is in rough agreement with the optical theorem, although we should not expect equation (12.67) to hold exactly because we have used the Born approximation.

We now check the validity of the Born approximation. The potential of equation (12.72) has a characteristic range a_0. When an electron with momentum $\sim \bar{\mathbf{p}}$ is aimed at the atom, it is within this range for a time of order $\delta t \simeq a_0 m/\bar{p}$. Averaged over that time, the potential it experiences is of order

$$\overline{V} \equiv \frac{1}{a_0^3}\int \mathrm{d}r\, r^2 V(r) = -\frac{e^2}{8\pi\epsilon_0 a_0} = -\mathcal{R}, \tag{12.77}$$

where we have used the definition (8.13b) of a_0 and $\mathcal{R}$ is the Rydberg constant (eq. 8.21). From the TDSE the fractional change that $\overline{V}$ effects in its ket during this interval is of order $\delta|\psi\rangle/|\psi\rangle \sim \overline{V}\delta t/\hbar$. We expect the Born approximation to be is valid if this fractional change is small, that is, provided

$$1 \gg \frac{a_0 m}{\bar{p}}\frac{|\overline{V}|}{\hbar} = \frac{\sqrt{\mathcal{R}m_\mathrm{e}/2}}{\bar{p}}. \tag{12.78}$$

Hence the inequality holds for electrons with energies

$$\frac{\bar{p}^2}{2m_\mathrm{e}} \gg \tfrac{1}{4}\mathcal{R}. \tag{12.79}$$

Since $\mathcal{R} \sim 13.6\,\mathrm{eV}$, while the rest-mass energy of the electron is $m_\mathrm{e}c^2 \sim 511\,\mathrm{keV}$, there is a wide range of energy that is high enough for the Born approximation to be valid, yet small enough for the electron to be non-relativistic. In Figure 12.4 we plot the experimentally measured differential cross section alongside our estimate (12.75) from the Born approximation for three electron energies: 4.9, 30 and 680 eV. At the lowest energy the Born approximation is useless. At $30\,\mathrm{eV} \sim 2\mathcal{R}$ the approximation works moderately well for back-scattering but seriously underpredicts the cross section for forward scattering. At 680 eV the approximation works well for all scattering angles.

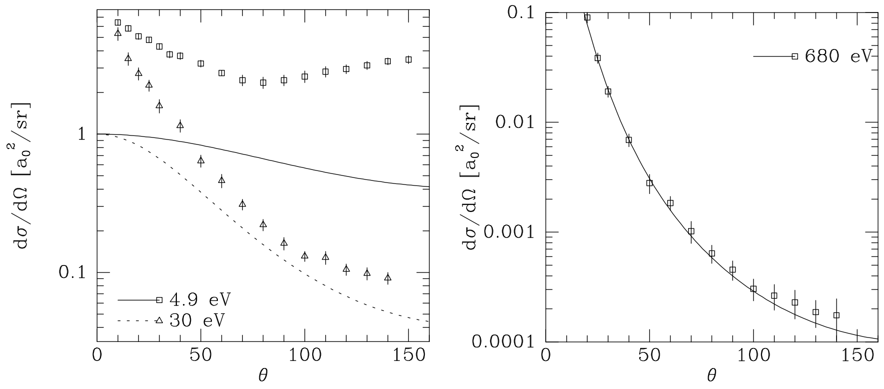


Figure 12.4 Elastic e^{-}H scattering at electron kinetic energies $E = 4.9$, 30 and 680 eV. The curves show the predictions of the Born approximation (eq. 12.75) while the points show experimental data from J.F. Williams, J. Phys. B, **8** no. 13 (1975). The accuracy of the Born approximation increases with energy.

12.5 Partial wave expansions

In §12.2 we introduced the S-matrix by squeezing the scattering operator between states $|\mathbf{p}\rangle$ of definite momentum. This allowed us to evaluate the action of the free evolution operators $U_{\mathrm{K}}(\tau)$, because $\{|\mathbf{p}\rangle\}$ is a complete set of eigenstates of H_{K}. In §7.2.5 we saw that states $|E, l, m\rangle$ of definite *angular* momentum also form a complete set of eigenstates of H_{K}, so we could just as well consider the matrix $\langle E', l', m'|\mathcal{S}|E, l, m\rangle$.

From equation (12.37) we have that $[H_{\mathrm{K}}, \mathcal{S}] = 0$, from which it follows that $\langle E', l', m'|\mathcal{S}|E, l, m\rangle$ vanishes unless $E' = E$. Also, if the scattering potential is spherically symmetric, it follows from the work of §4.2 that $[\mathbf{L}, \mathcal{S}] = 0$, so

$$[L_z, \mathcal{S}] = 0 \quad ; \quad [L_\pm, \mathcal{S}] = 0 \quad ; \quad [L^2, \mathcal{S}] = 0. \tag{12.80}$$

From the first and last of these commutators it follows that $\langle E', l', m'|\mathcal{S}|E, l, m\rangle$ vanishes unless $m' = m$ and $l' = l$. Moreover, the second commutator implies that[6]

$$\begin{aligned} 0 &= \langle E, l, m|[\mathcal{S}, L_+]|E, l, m-1\rangle \\ &\propto \langle E, l, m|\mathcal{S}|E, l, m\rangle - \langle E, l, m-1|\mathcal{S}|E, l, m-1\rangle, \end{aligned} \tag{12.81}$$

so not only is $\mathcal{S}$ diagonal in the $|E, l, m\rangle$ basis, but $\langle E, l, m|\mathcal{S}|E, l, m\rangle$ is actually independent of m. We can summarise these constraints on the S-matrix of a spherically symmetric potential by writing

$$\langle E', l', m'|\mathcal{S}|E, l, m\rangle = \delta(E - E')\delta_{l'l}\delta_{m'm}\, s_l(E), \tag{12.82}$$

where $s_l(E)$ is a number that depends on E and l. Finally, since the S-matrix is unitary, $s_l(E)$, must have unit modulus, so

$$\langle E', l', m'|\mathcal{S}|E, l, m\rangle = \delta(E - E')\delta_{l'l}\delta_{m'm}\, \mathrm{e}^{2\mathrm{i}\delta_l(E)}, \tag{12.83}$$

where all the remaining information is contained in the real **phase shifts** $\delta_l(E)$. This reduction of the whole scattering process to a mere set of phases makes the angular-momentum basis invaluable for scattering problems.

Equation (12.83) implies that

$$\mathcal{S}|E, l, m\rangle = \mathrm{e}^{2\mathrm{i}\delta_l(E)}|E, l, m\rangle, \tag{12.84}$$

---

[6] Recall from equations (7.15) that $\alpha_+(m-1) = \alpha_-(m)$.

Box 12.1: The amplitude $\langle\mathbf{p}|E,l,m\rangle$

We require the amplitude $\langle\mathbf{p}|E,l,m\rangle$ that a particle that has well-defined energy E and angular momentum will be found to have momentum $\mathbf{p}$ and energy $p^2/2M$. Since the quantity we seek is the momentum-space wavefunction of a particle of well-defined angular momentum, we simply repeat the work of §7.2.3 in the momentum representation. We have

$$L_z = \frac{1}{\hbar}(p_y x - p_x y) = \mathrm{i}\left(p_y\frac{\partial}{\partial p_x} - p_x\frac{\partial}{\partial p_y}\right),$$

where the operators are written in the momentum representation. We now introduce polar coordinates (p,ϑ,φ) for momentum space, and, in exact analogy to the derivation of equation (7.43), show that $L_z = \mathrm{i}\partial/\partial\varphi$. This is simply minus the corresponding real-space result (7.43). It is easy to see that proceeding in this way we would obtain momentum-space representations of $L_\pm$ that differ from their real-space analogues (7.52) only in the substitutions $\theta \to \vartheta$ and $\phi \to \varphi$ and an overall change of sign. Consequently the momentum-space wavefunction of a state of well defined angular momentum must be $\psi(\mathbf{p}) = g(p)\mathrm{Y}_l^{m*}(\vartheta,\varphi)$, where $g(p)$ is as yet undetermined and the complex conjugate spherical harmonic is required because $L_z = +\mathrm{i}\partial/\partial\varphi$. If we require E to equal $p^2/2M$, it is clear that $g = G\delta(E - p^2/2M)$. The constant of proportionality, G, is determined by the normalization condition

$$\begin{aligned}\delta(E-E') &= \langle E,l,m|E',l,lm\rangle\\ &= G^2\int \mathrm{d}p\, p^2\delta(E-p^2/2M)\delta(E'-p^2/2M)\int \mathrm{d}\Omega|\mathrm{Y}_l^m|^2\\ &= G^2M\int \mathrm{d}E_p\,\sqrt{2ME_p}\delta(E-E_p)\delta(E'-E_p)\\ &= G^2M\sqrt{2ME}\,\delta(E-E') = G^2Mp\delta(E-E').\end{aligned}$$

Thus $G = (Mp)^{-1/2}$.

so if prior to scattering the particle is in the state $|E,l,m\rangle$, it will emerge from the scattering region in a state that differs only by the acquisition of an extra phase $2\delta_l(E)$.

In fact, the above discussion generalises straightforwardly to the case of particles with non-zero spin, provided we replace the orbital angular momentum operator $\mathbf{L}$ with the total angular momentum operator $\mathbf{J}$, and relabel the states and phase shifts accordingly. For simplicity, we will confine ourselves to scalar particles for the rest of this section.

To relate the S-matrix in the form of equation (12.83) to experimental cross-sections, we must calculate the scattering amplitude $f(\mathbf{p}\to\mathbf{p}')$. Using equations (12.6) and (12.83) we obtain

$$\begin{aligned}\langle\mathbf{p}'|T|\mathbf{p}\rangle &= \sum_{l'lm'm}\int \mathrm{d}E\mathrm{d}E'\,\langle\mathbf{p}'|E',l',m'\rangle\langle E',l',m'|T|E,l,m\rangle\langle E,l,m|\mathbf{p}\rangle\\ &= \sum_{lm}\int \mathrm{d}E\,\langle\mathbf{p}'|E,l,m\rangle\langle E,l,m|\mathbf{p}\rangle(\mathrm{e}^{2\mathrm{i}\delta_l(E)}-1).\end{aligned} \tag{12.85}$$

In Box 12.1 we show that

$$\langle\mathbf{p}|E,l,m\rangle = (Mp)^{-1/2}\delta(E-E_p)\mathrm{Y}_l^{m*}(\vartheta,\varphi), \tag{12.86}$$

where (ϑ,φ) are the polar coordinates of $\mathbf{p}$.

If we align the z-axis with the beam direction, the initial state is unchanged by rotations around the z-axis. Equation (7.38) with $\boldsymbol{\alpha} = \hat{\mathbf{z}}$ then tells that the initial state has $m = 0$, so $\langle p\hat{\mathbf{z}}|E, l, m\rangle$ vanishes unless $m = 0$. Thus for a beam in this direction

$$\langle E, l, m|\mathbf{p}\rangle = \frac{\delta(E_p - E)}{\sqrt{Mp}}\,\delta_{m0}\mathrm{Y}_l^0(0). \tag{12.87}$$

In §7.2.3 we saw that $\mathrm{Y}_l^0(\vartheta)$ is a real l^{th}-order polynomial in $\cos\vartheta$:

$$\mathrm{Y}_l^0(\vartheta) = \sqrt{\frac{2l+1}{4\pi}}\,\mathrm{P}_l(\cos\vartheta). \tag{12.88}$$

We have, moreover, that $\mathrm{P}_l(1) = 1$. Consequently, equation (12.87) yields

$$\langle E, l, m|\mathbf{p}\rangle = \delta(E_p - E)\delta_{m0}\sqrt{\frac{2l+1}{4\pi Mp}}. \tag{12.89}$$

Using equation (12.86) again to eliminate $\langle \mathbf{p}'|E, l, m\rangle$ from equation (12.85), and then using equation (12.88) to eliminate Y_l^0, we obtain

$$\begin{aligned}\langle \mathbf{p}'|T|\mathbf{p}\rangle &= \sum_l \int \mathrm{d}E\,\langle \mathbf{p}'|E, l, 0\rangle\langle E, l, 0|\mathbf{p}\rangle\big(\mathrm{e}^{2\mathrm{i}\delta_l(E_p)} - 1\big)\\ &= \sum_l \frac{2l+1}{4\pi M}\int \frac{\mathrm{d}E}{\sqrt{p'p}}\,\delta(E - E_{p'})\delta(E_p - E)\mathrm{P}_l(\cos\vartheta')\big(\mathrm{e}^{2\mathrm{i}\delta_l(E_p)} - 1\big)\\ &= \frac{\mathrm{i}}{2\pi\hbar M}\delta(E_p - E_{p'})\sum_l \frac{2l+1}{2\mathrm{i}p}\hbar\mathrm{P}_l(\cos\vartheta')\big(\mathrm{e}^{2\mathrm{i}\delta_l(E)} - 1\big),\end{aligned} \tag{12.90}$$

where ϑ' is the angle between $\mathbf{p}$ and $\mathbf{p}'$. Comparing this equation with the definition (12.39) of the scattering amplitude, we see finally that

$$f(\mathbf{p} \to \mathbf{p}') = \sum_l (2l+1)\mathrm{P}_l(\cos\vartheta')f_l(E_p), \tag{12.91a}$$

where the **partial-wave amplitude** is defined to be

$$f_l(E) \equiv \hbar\frac{\mathrm{e}^{2\mathrm{i}\delta_l(E)} - 1}{2\mathrm{i}p} = \frac{\hbar}{p}\,\mathrm{e}^{\mathrm{i}\delta_l(E)}\sin\delta_l(E). \tag{12.91b}$$

The differential cross-section is just the mod square of the scattering amplitude, and because the spherical harmonics are orthonormal when integrated over all angles, the total cross-section is

$$\begin{aligned}\sigma_{\text{tot}} &= \int \mathrm{d}\Omega\,|f(\mathbf{p} \to \mathbf{p}')|^2\\ &= 4\pi\sum_{l'l}\sqrt{(2l'+1)(2l+1)}f_{l'}(E)^* f_l(E)\int \mathrm{d}\Omega\,\mathrm{Y}_{l'}^0(\vartheta')\mathrm{Y}_l^0(\vartheta')\\ &= 4\pi\sum_l (2l+1)|f_l(E)|^2 = 4\pi\hbar^2\sum_l \frac{2l+1}{p^2}\sin^2\delta_l(E).\end{aligned} \tag{12.92}$$

This equation is often written as $\sigma_{\text{tot}} = \sum_l \sigma_l$, where the **partial cross-section** of order l,

$$\sigma_l \equiv 4\pi(2l+1)|f_l(E)|^2 = 4\pi\hbar^2\frac{2l+1}{p^2}\sin^2\delta_l(E), \tag{12.93}$$

is the cross-section for scattering a particle that has total squared angular momentum $l(l+1)\hbar^2$. Clearly, the partial cross-sections are restricted by

$$0 \le \sigma_l \le 4\pi\hbar^2\frac{2l+1}{p^2}, \tag{12.94}$$

with σ_l only vanishing when the phase shift $\delta_l = n\pi$. Notice from equations (12.91a) that

$$\Im\mathrm{m} f(\mathbf{p} \to \mathbf{p}) = \sum_l (2l+1)\mathrm{P}_l(1)\Im\mathrm{m} f_l(E) = \sum_l \frac{2l+1}{p}\hbar \sin^2 \delta_l(E). \quad (12.95)$$

Comparison of this with equation (12.92) shows that the optical theorem (12.67) is explicitly satisfied in this basis. This fact follows from conservation of angular momentum – we have treated the incoming beam as a superposition of states of well-defined angular momentum; since the potential is spherically symmetric, it cannot change the particle's angular momentum, so each state of well-defined angular momentum scatters separately, and does so in conformity with the optical theorem.

In the classical picture of scattering, the angular momentum of a particle of energy $p^2/2M$ is determined by its impact parameter b, which is the distance between the scattering centre and the straight line tangent to the incoming trajectory. Quantitatively, the angular momentum has magnitude $L = bp$. Large b corresponds to a glancing collision and a small scattering angle, while at small b the encounter is nearly head-on and the particle is liable to back-scatter. Thus we expect the differential cross section $f_l(\mathbf{p} \to \mathbf{p}')$ to be largest for $\vartheta' \approx 0$ when l is large, and smallest for $\vartheta' \approx \pi$ when $l \approx 0$. The total cross section σ_l is expected to decrease as l, and therefore b, increases.

12.5.1 Scattering at low energy

At low energy, p is small and for $l > 0$ the classical impact parameter $b = L/p$ becomes large. Hence we expect low-energy scattering to be dominated by the partial wave with $l = 0$. In this subsection we show that this naive expectation is borne out by our quantum-mechanical formulae.

To discover how a given particle is actually scattered, we must relate the phase shifts $\delta_l(E)$ to the scattering potential $V(r)$. Since the free state $|E, l, m\rangle$ is an eigenstate of H_K, L^2 and L_z, from equation (7.70) it follows that[7]

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\langle \mathbf{r}|E, l, m\rangle\right) = \left(\frac{l(l+1)}{r^2} - \frac{2mE}{\hbar^2}\right)\langle \mathbf{r}|E, l, m\rangle. \quad (12.96)$$

When $l \neq 0$, the angular momentum term dominates the right side near the origin, and one can easily show that $\langle \mathbf{r}|E, l, m\rangle \sim r^l$ for small r (Problem 12.8). Consequently, there is only a very small probability of finding a particle that has high angular momentum near the origin. This reasoning breaks down when the second term on the right side becomes important. For energy $E = p^2/2m$ this occurs at $r \approx l\hbar/p$, which for large l coincides with the classical impact parameter. Suppose that the scattering potential acts over some characteristic length R, beyond which it is negligible – for example, in the case of the potential (12.72), R is of the order of a few Bohr radii. If $R \ll l\hbar/p$, then $V(r)$ is only strong in a region where a free wavefunction would be very small. In this case, the l^{th} partial wave will scarcely be affected, so $\delta_l \simeq 0$. Roughly, the only states that suffer significant scattering are those with angular momenta in the range $l\hbar \lesssim pR$, so for low incoming momenta the total scattering amplitude can be well approximated by the first few terms in the infinite sum (12.91a).

In fact, we can see quite generally that only the lowest l states make significant contributions to the low-energy cross-section. The scattering amplitude $f(\mathbf{p} \to \mathbf{p}')$ should be a smooth function of $\mathbf{p}$ and $\mathbf{p}'$ as $p, p' \to 0$,

---

[7] In fact, $\langle \mathbf{r}|E, l, m\rangle = j_l(kr)\mathrm{Y}_l^m(\theta, \phi)$, where $k = \sqrt{2ME}/\hbar$ and j_l is the l^{th} spherical Bessel function, but we do not need this result here (see Problem 12.8).

because at low energies the incoming particle has a large wavelength and cannot resolve any sharp features in the potential. Since in equation (12.91a) P_l is an l^{th}-order polynomial in $\cos\vartheta = \mathbf{p}'\cdot\mathbf{p}/p'p$, we see that if $f(\mathbf{p}\to\mathbf{p}')$ is to be an analytic function of the Cartesian components of $\mathbf{p}$ and $\mathbf{p}'$ at low energies, the partial wave amplitude $f_l(E)$ must vanish with p at least as fast as

$$\lim_{p\to 0} f_l(E) \sim (p'p)^l = p^{2l}. \tag{12.97}$$

The total cross section (12.92) then behaves as

$$\lim_{p\to 0} \sigma_{\text{tot}} = 4\pi\sum_l a_l^2 p^{4l} \tag{12.98}$$

in terms of some constants a_l, and can be well approximated by just the lowest few terms. In the extreme low energy limit, the only non-vanishing amplitude is $f_0(E_p\to 0) = a_0$ and the differential cross section

$$\lim_{p\to 0} \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq a_0^2\mathrm{P}_0 = a_0^2 \tag{12.99}$$

is isotropic.

An eigenstate of the true Hamiltonian H with the same energy and angular-momentum quantum numbers as $|E,l,m\rangle$ has a radial wavefunction $u_l(r)$ that satisfies a version of equation (12.96) modified by the inclusion of a potential $V(r)$. Writing $u_l(r) = U_l(r)/r$ we find

$$\frac{\mathrm{d}^2U_l(r)}{\mathrm{d}r^2} = -\frac{2mE}{\hbar^2}U_l(r) + V_{\text{eff}}U_l(r), \tag{12.100a}$$

where

$$V_{\text{eff}}(r) \equiv \frac{2mV(r)}{\hbar^2} + \frac{l(l+1)}{r^2}. \tag{12.100b}$$

For a general potential, we typically have to solve this equation numerically, and then find the phase shifts by comparing our solution with equation (12.45) in the large r limit. However, we can obtain a heuristic understanding of the behaviour of the phase shift as follows. If the potential is attractive ($V(r) < 0$), $U_l(r)$ will have a greater curvature, and hence oscillate more rapidly in the presence of V that it would have done if the particle were free. A potential with finite range $R \ll l\hbar/p$ only acts over a small part of a radial oscillation, so when $V < 0$, the wavefunction emerges from the interaction region slightly further along its cycle than a free wavefunction. On the other hand, when the potential is repulsive, $U_l(r)$ has smaller curvature, so oscillates more slowly than a free wavefunction, emerging from the interaction region slightly behind. Equation (12.84) tells us that states emerge from the scattering process changed only in phase; we now see that the sign of the phase shift δ_l will typically be opposite to that of the potential.

As the magnitude of V increases, so does the difference in oscillation rates between interacting and free eigenstates, and hence at fixed energy $|\delta_l(E)|$ likewise increases. When the potential is sufficiently strong, the interacting wavefunction can oscillate precisely half a cycle more (or less if $V > 0$) in the interaction region than it would do if the state were free, and then $|\delta_l(E)| = \pi$ with the consequence that $f_l(E) \to 0$. In these circumstances this angular-momentum state suffers no scattering at all.

In §10.3, we saw that atoms of a noble gas such as argon are chemically inert because in their ground states they have spherically-symmetric distributions of electron charge, and they have no low-lying excited states that can be mixed in by a perturbation to generate a less symmetrical charge distribution. As a consequence, these atoms generate negligible electric fields beyond some limiting radius $R \simeq a_0$ that contains nearly all the probability

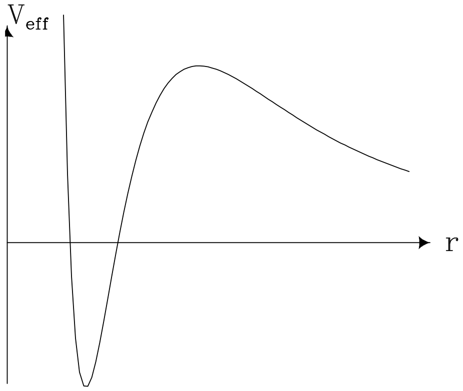


Figure 12.5 An exponential attractive force combined with centrifugal repulsion generates a minimum in the effective potential.

density of the outermost shell of electrons. At $r \leq R$ there is a significant electric field, and any particle that penetrates this region will be appreciably scattered, but particles that have larger impact parameters will be negligibly scattered. At energies low enough that $Rp \ll \hbar$, scattering from states with $l > 0$ is negligible, while the considerations of the previous paragraph suggest that there could be an energy at which there is also no scattering from states with $l = 0$. Then an electron is not scattered at all. Exactly this **Ramsauer–Townsend effect** was observed before the development of quantum mechanics[8] when electrons of energy $\sim 0.7\,\mathrm{eV}$ were discovered to pass undeflected past noble-gas atoms.

12.6 Resonances

Physically interesting objects are usually sufficiently complicated that the way they scatter approaching particles has to be studied numerically. However, we can sometimes see that all potentials that have some generic feature will scatter particles in a particular way. For example, the effective potential V_{eff} of equation (12.100b) is likely to have a local minimum if the true potential V falls off with r more steeply than r^{-2} as $r \to \infty$ (see Figure 12.5). A particle thrown in from a large distance can, if it has roughly the right energy, get trapped for a while in a local minimum, before eventually escaping to the true ground state at infinity. Therefore, even if a target's potential is unknown, we can infer that it contains such dips if we notice that a particle fired towards the target is delayed before re-emerging.

Equation (12.46) gives the wavefunction at late times of a particle that was initially in the free state $|\phi\rangle$. It breaks the wavefunction into two parts. The first is a sum of plane waves $\phi(\mathbf{p})\mathrm{e}^{\mathrm{i}(\mathbf{p}\cdot\mathbf{r}-Et)/\hbar}$. If $|\phi(\mathbf{p})|^2$ has a well-defined peak at momentum $\bar{\mathbf{p}}$, from §2.3.3 we know that the amplitude of this first contribution peaks on the plane $\hat{\bar{\mathbf{p}}}\cdot\mathbf{r} = \bar{p}t/m$. To determine the location at which the amplitude of the second part peaks, we observe that as $t \to \infty$, the phase of the exponential in equation (12.46) varies extremely rapidly as a function of momentum, causing the sign of the integrand to oscillate quickly. If $\phi(\mathbf{p})$ is a smooth function, the integral will be dominated by the contribution from momenta at which the phase is stationary. To find these points, we must take into account the phase of $f(\mathbf{p} \to p\hat{\mathbf{r}})$. By equations (12.91), the scattering amplitude for each partial wave is real except for a factor $\mathrm{e}^{\mathrm{i}\delta_l(E_p)}$. Hence the dominant contribution to the second term in equation (12.46) arises when

$$\frac{\partial}{\partial p}\left\{pr - E_p t + \hbar\delta_l(E_p)\right\} = 0 \quad \text{i.e. when} \quad r = \frac{p}{m}t - \hbar\frac{\partial}{\partial p}\delta_l(E_p). \qquad (12.101)$$

---

[8] C. Ramsauer, Ann. Physik, **4**, 64 (1921); V.A. Bailey & J.S. Townsend, Phil. Mag., **42**, 873 (1921).

We see that if the phase shift $\delta_l(E_p)$ increases sharply for momenta near the average momentum of the initial state, the amplitude of the wave will be concentrated at a smaller radius than the incoming wave would have been. Consequently, there will be two distinct peaks in the probability of a particle reaching a detector at some distance from the scattering centre. The first is associated with the possibility that the particle misses its target, and the second with the possibility that it hits and is temporarily trapped by it before making good its escape. Thus unstable bound states are associated with rapid increases in the phase shift of the scattered particle.

We can model a dramatic increase of $\delta_l(E)$ by postulating that, for energy near some value E_{R}, the phase shift behaves as

$$\delta_l(E) \simeq \tan^{-1}\left(\frac{-\Gamma/2}{E - E_{\mathrm{R}}}\right), \tag{12.102}$$

where the fixed energy scale Γ is included to ensure that the argument of the inverse tangent is dimensionless, and must be positive if we want $\partial\delta_l/\partial p > 0$. In this model the phase shift rapidly increases by π as the energy increases through E_{R}. Using the model in equation (12.101), we find that the time delay between the two peaks in the probability density of the particle hitting a given detector is

$$\begin{aligned}\Delta t = \frac{m\hbar}{p}\frac{\partial}{\partial p}\delta_l(E) &= \frac{m\hbar}{p}\frac{\partial E}{\partial p}\frac{\partial}{\partial E}\tan^{-1}\left(\frac{-\Gamma/2}{E - E_{\mathrm{R}}}\right)\\ &= \frac{\hbar\Gamma/2}{(E - E_{\mathrm{R}})^2 + (\Gamma/2)^2}.\end{aligned} \tag{12.103}$$

We infer from this delay that the lifetime of the quasi-bound state is $\approx \hbar/\Gamma$.

Calculating Γ for a physically realistic potential usually requires numerical analysis. However, since the lifetime of the quasi-bound state increases as Γ decreases, we anticipate that smaller values of Γ correspond to deeper minima in the potential: a deeper well traps the particle for longer. The limiting case $\Gamma \to 0^+$ implies that the delay in emergence becomes infinite. We interpret this to mean that the dip in V is just deep enough to genuinely bind an incoming particle.

If V is so deep that there is a state that has a strictly negative energy, the final state may not resemble the initial free state. For example, the incoming particle may get trapped for good in the potential, or it may knock out another particle that is already trapped (as in ionisation of an atom). In such cases, the scattering is said to be **inelastic** and the methods of this chapter must be extended[9].

Returning to the case $\Gamma > 0$, we now investigate how the cross-section is affected by the delayed emergence of our particle. From equations (12.6) and (12.46) the wavefunction of the scattered particle in the asymptotic future is

$$\begin{aligned}\langle \mathbf{r}|\mathcal{T}|\phi;t\rangle &= \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi\hbar)^{3/2}}\phi(\mathbf{p})\frac{\mathrm{e}^{\mathrm{i}(pr-Et)/\hbar}}{r}f(\mathbf{p}\to p\hat{\mathbf{r}})\\ &= \sum_l (2l+1)\int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi\hbar)^{3/2}}\phi(\mathbf{p})\frac{\mathrm{e}^{\mathrm{i}(pr-Et)/\hbar}}{r}\mathrm{P}_l(\hat{\mathbf{p}}\cdot\hat{\mathbf{r}})f_l(E),\end{aligned} \tag{12.104}$$

where the second line uses (12.91)to relate $f(\mathbf{p}\to\mathbf{p}')$ to the partial-wave amplitudes $f_l(E)$. To derive the cross-section (12.59), we assumed that the scattering amplitudes are more slowly varying functions of $\mathbf{p}$ than the wavepacket

---

[9] The difficulty is not too severe. True bound states have energy $E < 0$ whereas all free states must have energy $E \geq 0$. Hence, if $|\mathrm{b}\rangle$ is bound and $|\phi\rangle$ is free, $\langle \mathrm{b}|\phi\rangle = 0$ so H acts on a larger Hilbert space than does H_{K}. Including these extra states carefully allows us to treat bound states. (See also Problem 12.6.)

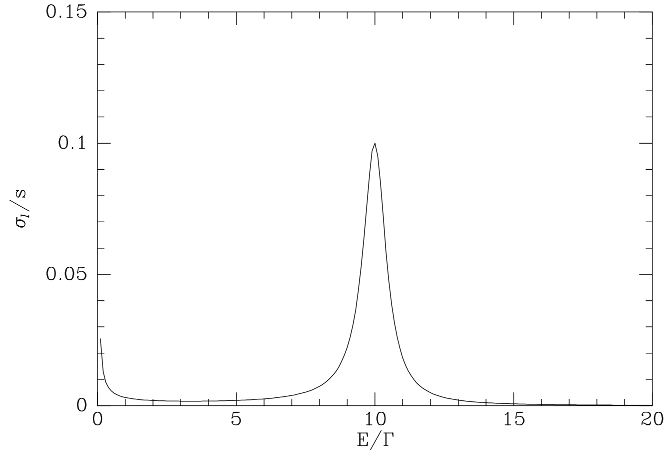


Figure 12.6 The Breit–Wigner formula for a scattering cross section in the presence of a resonance. Here $E_{\rm R} = 10\Gamma$ and the cross section is normalised by $s \equiv 2(2l+1)\hbar^2/m\Gamma$.

$\phi(\mathbf{p})$ – see the discussion after equation (12.54). In the presence of a resonance, this approximation may break down. Indeed, when one of the phase shifts $\delta_l(E)$ has the form of equation (12.102), from equation (12.91b) the corresponding partial wave amplitude is

$$f_l(E) \equiv \frac{\hbar}{p}\,{\rm e}^{{\rm i}\delta_l(E)} \sin\delta_l(E) = -\frac{\hbar}{p}\frac{\Gamma/2}{E - E_{\rm R} + {\rm i}\Gamma/2}, \qquad (12.105)$$

and for small Γ this varies rapidly when $E \approx E_{\rm R}$. We assume that the incoming wavepacket $\phi(\mathbf{p})$ contains states $|\mathbf{p}\rangle$ that are restricted in energy to a range of width Δ and consider first the case $\Delta \ll \Gamma$ in which the resonance is broader than the uncertainty in the energy of the incoming particle.

12.6.1 Breit–Wigner resonances

If $\Delta \ll \Gamma$, then $f_l(E)$ varies slowly with energy in comparison to $\phi(\mathbf{p})$, and equation (12.93) for the partial cross-sections σ_l applies. Again using equation (12.102), we have

$$\sigma_l(E) = 4\pi(2l+1)|f_l|^2 \simeq \frac{4\pi\hbar^2}{p^2}\frac{(2l+1)(\Gamma/2)^2}{(E-E_{\rm R})^2 + (\Gamma/2)^2}. \qquad (12.106)$$

A peak in the cross-section that follows this famous formula is said arise from a pure **Breit–Wigner resonance**. Breit–Wigner resonances are easily detected in plots of a cross-section versus energy and are the experimental signature of quasi-bound states in the scattering potential. Figure 12.6 is a plot of equation (12.106). Notice that the energy dependence is a combination of the slow decline with E associated with the factor p^{-2} and the peak that arises from the Lorentzian final factor – such factors are familiar from the theory of a damped harmonic oscillator. If $\Gamma \ll E_{\rm R}$, the factor p^{-2} changes very little over the width of the bump, and the resonance curve falls to half its maximum height when $E \simeq E_{\rm R} \pm \Gamma/2$. Thus the resonance lifetime $\hbar/\Gamma$ can be determined from the FWHM of the peak in the cross-section. This result explains why we needed to restrict ourselves to superpositions with $\Delta \ll \Gamma$: in order to resolve the Breit–Wigner curve experimentally, there had better be a good chance that our particle's energy lies within Γ of $E_{\rm R}$, where all the action lies.

12.6.2 Radioactive decay

The width Γ of a very long-lived resonance may be so small that our experimental apparatus cannot generate incoming particles with sufficiently small uncertainty Δ in the energy to resolve the curve of Figure 12.6. Then, using equation (12.86), we expand the momentum amplitudes $\phi(\mathbf{p})$ of the initial state as

$$\begin{aligned}\phi(\mathbf{p}) &= \sum_{lm}\int \mathrm{d}E'\,\langle\mathbf{p}|E',l,m\rangle\langle E',l,m|\phi\rangle\\ &= \sum_{lm} \mathrm{Y}_l^{m*}(\hat{\mathbf{p}})\frac{\langle E,l,m|\phi\rangle}{\sqrt{Mp}}, \quad \text{where} \quad E\equiv\frac{p^2}{2M}.\end{aligned} \tag{12.107}$$

Suppose for simplicity that, near some energy E_{R}, only one partial wave amplitude has the form of equation (12.105), the others all being negligible by comparison. Then, ignoring any angular dependence, the final-state wavefunction (12.104) of the scattered particle contains a factor

$$\begin{aligned}\langle\mathbf{r}|\mathcal{T}|\phi;t\rangle &\propto \int \mathrm{d}p\,\frac{p^2}{(2\pi\hbar)^{3/2}}\frac{\langle E,l,m|\phi\rangle}{(Mp)^{1/2}}\frac{\Gamma\hbar}{2pr}\frac{\mathrm{e}^{\mathrm{i}(pr-Et)/\hbar}}{E-E_{\mathrm{R}}+\mathrm{i}\Gamma/2}\\ &= \frac{\Gamma\hbar}{2r}\int_0^\infty \frac{\mathrm{d}E}{(2\pi\hbar)^{3/2}}\sqrt{\frac{M}{p}}\langle E,l,m|\phi\rangle\frac{\mathrm{e}^{\mathrm{i}(pr-Et)/\hbar}}{E-E_{\mathrm{R}}+\mathrm{i}\Gamma/2}.\end{aligned} \tag{12.108}$$

If the initial state $|\phi\rangle$ has average energy around $\langle H_{\mathrm{K}}\rangle \simeq E_{\mathrm{R}}$, but is a superposition of states with different energies, smooth over a range $\Delta \gg \Gamma$, we may approximate $p^{-1/2}\langle E,l,m|\phi\rangle$ by its value at resonance, $p_{\mathrm{R}}^{-1/2}\langle E_{\mathrm{R}},l,m|\phi\rangle$ and bring it outside the integral. Similarly, near resonance we approximate p by

$$p \simeq p_{\mathrm{R}} + \frac{\mathrm{d}p}{\mathrm{d}E}(E-E_{\mathrm{R}}) = p_{\mathrm{R}} + \frac{E-E_{\mathrm{R}}}{v_{\mathrm{R}}}, \tag{12.109}$$

where $v_{\mathrm{R}} = p_{\mathrm{R}}/M$, so

$$\mathrm{e}^{\mathrm{i}(pr-Et)/\hbar} \simeq \mathrm{e}^{\mathrm{i}(p_{\mathrm{R}}r-E_{\mathrm{R}}t)/\hbar}\mathrm{e}^{\mathrm{i}(E-E_{\mathrm{R}})(r/v_{\mathrm{R}}-t)/\hbar}. \tag{12.110}$$

Substituting these approximations into equation (12.108), we find

$$\langle\mathbf{r}|\mathcal{T}|\phi;t\rangle \sim \frac{\Gamma\hbar}{2v_{\mathrm{R}}^{1/2}}\frac{\langle E_{\mathrm{R}},l,m|\phi\rangle}{(2\pi\hbar)^{3/2}}\frac{\mathrm{e}^{\mathrm{i}(p_{\mathrm{R}}r-E_{\mathrm{R}}t)/\hbar}}{r}\int_0^\infty \mathrm{d}E\,\frac{\mathrm{e}^{\mathrm{i}(E-E_{\mathrm{R}})(r/v_{\mathrm{R}}-t)/\hbar}}{E-E_{\mathrm{R}}+\mathrm{i}\Gamma/2}. \tag{12.111}$$

The remaining integral can be done by contour integration. Since the denominator is large except near $E = E_{\mathrm{R}}$, we can extend the range of integration to $-\infty < E < \infty$, without drastically affecting the integral. If $r > v_{\mathrm{R}}t$, we close the contour in the upper half complex E plane. Since the only pole is in the lower half-plane, the integral evaluates to zero. If $r < v_{\mathrm{R}}t$, we close the contour in the lower half-plane. Evaluating the residue of the pole at $E = E_R - \mathrm{i}\Gamma/2$, we conclude that

$$\langle\mathbf{r}|\mathcal{T}|\phi;t\rangle \sim \frac{\mathrm{i}\Gamma}{2(2\pi\hbar v_{\mathrm{R}})^{1/2}}\langle E_{\mathrm{R}},l,m|\phi\rangle\frac{\mathrm{e}^{\mathrm{i}(p_{\mathrm{R}}r-E_{\mathrm{R}}t)/\hbar}}{r}\mathrm{e}^{-\Gamma(t-r/v_{\mathrm{R}})/2\hbar}. \tag{12.112}$$

Consequently,

$$|\langle\mathbf{r}|\mathcal{T}|\phi;t\rangle|^2 \sim \begin{cases} 0 & \text{if } r > v_{\mathrm{R}}t \\ \dfrac{\Gamma^2}{8\pi\hbar v_{\mathrm{R}}}\dfrac{|\langle E_{\mathrm{R}},l,m|\phi\rangle|^2}{r^2}\,\mathrm{e}^{-\Gamma(t-r/v_{\mathrm{R}})/\hbar} & \text{otherwise.}\end{cases} \tag{12.113}$$

The physical interpretation of this equation is the following. The probability density $|\langle \mathbf{r}|T|\phi;t\rangle|^2$ is zero before time $t' = r/v_{\rm R}$ because the particle travels radially outwards at speed $\simeq v_{\rm R}$. Subsequently, the probability of finding the particle anywhere on a sphere of radius r decays exponentially as ${\rm e}^{-\Gamma(t-t')/\hbar}$.

This result provides a remarkable explanation of the law of radioactive decay: we interpret the emission of a neutron by an unstable nucleus as the endpoint of a scattering experiment that started months earlier in a nuclear reactor, where the nucleus was created by absorption of a neutron. More dramatic is the case of ^{238}U, which decays via emission of an α-particle to ^{234}Th with a mean life $\hbar/\Gamma \simeq 6.4\,$Gyr. Because $\Gamma/\hbar$ is tiny, the probability (12.113) is nearly constant over huge periods of time. Our formalism tells us that if we were to scatter α-particles off ^{234}Th, they would all eventually re-emerge, but only after a delay that often exceeds the age of the universe! Thus ^{238}U is really a long-lived resonance of the $(\alpha,^{234}{\rm Th})$ system, rather than a stationary state. It is only because the timescale $\hbar/\Gamma$ is so long that we speak of ^{238}U rather than a resonance in the (α, ^{234}Th) system. In fact, ^{234}Th is itself a resonance, ultimately of Pb. The longevity of ^{238}U is inevitably associated with a very small probability that ^{238}U will be formed when we shoot an α particle at a ^{234}Th nucleus. To see this notice that the final-state wavefunction $\langle \mathbf{r}|\mathcal{S}|\phi;t\rangle = \langle \mathbf{r}|\phi;t\rangle + \langle \mathbf{r}|\mathcal{T}|\phi;t\rangle$, also involves an unscattered piece. On account of the smallness of Γ, the ratio of probabilities

$$\frac{\text{Prob}(\alpha \text{ is trapped})}{\text{Prob}(\alpha \text{ unscattered})} \approx \frac{\Gamma^2 m}{\hbar p_{\rm R}}|\langle E_{\rm R}, l, m|\phi\rangle|^2 \tag{12.114}$$

is extremely small. Hence it is exceptionally difficult to form ^{238}U by firing α particles at ^{234}Th nuclei. Naturally occurring ^{238}U was formed in supernovae, where the flux of α-particles and neutrons was large enough to overcome this suppression.

Problems

12.1 Show that the operators $\widetilde{\Omega}_\pm$ defined by equation (12.23) obey

$$H\widetilde{\Omega}_\pm = \widetilde{\Omega}_\pm(H_{\rm K} \pm {\rm i}\epsilon) \mp {\rm i}\epsilon. \tag{12.115}$$

12.2 Obtain the first and second order contributions to the S-matrix from the Feynman rules given in §12.3.

12.3 Derive the **Lippmann–Schwinger equation**

$$|\pm\rangle = |E\rangle + \frac{1}{E - H_{\rm K} \pm {\rm i}\epsilon} V|\pm\rangle, \tag{12.116}$$

where $|\pm\rangle$ are in and out states of energy E and $|E\rangle$ is a free-particle state of the same energy. In the case that the potential $V = V_0|\chi\rangle\langle\chi|$ for some state $|\chi\rangle$ and constant V_0, solve the Lippmann–Schwinger equation to find $\langle\chi|\pm\rangle$.

12.4 A certain potential $V(r)$ falls as r^{-n} at large distances. Show that the Born approximation to the total cross-section is finite only if $n > 2$. Is this a problem with the Born approximation?

12.5 Compute the differential cross section in the Born approximation for the potential $V(\mathbf{r}) = V_0 \exp(-r^2/2r_0^2)$. For what energies is the Born approximation justified?

12.6 When an electron scatters off an atom, the atom may be excited (or even ionised). Consider an electron scattering off a hydrogen atom. The Hamiltonian may be written as $H = H_0 + H_1$ where

$$H_0 = \frac{\hat{\mathbf{p}}_1^2}{2m} - \frac{e^2}{4\pi\epsilon_0 r_1} + \frac{\hat{\mathbf{p}}_2^2}{2m} \tag{12.117}$$

is the Hamiltonian of the hydrogen atom (whose electron is described by coordinate $\mathbf{r}_1$) together with the kinetic Hamiltonian of the scattering electron (coordinate $\mathbf{r}_2$), and

$$H_1 = \frac{e^2}{4\pi\epsilon_0}\left(\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} - \frac{1}{r_2}\right) \tag{12.118}$$

is the interaction of the scattering electron with the atom.

By using H_0 in the evolution operators, show that in the Born approximation the amplitude for a collision to scatter the electron from momentum $\mathbf{p}_2$ to $\mathbf{p}_2'$ whilst exciting the atom from the state $|n, l, m\rangle$ to the state $|n', l', m'\rangle$ is

$$\begin{aligned} &f(\mathbf{p}_2; n, l, m \to \mathbf{p}_2'; n', l', m') \\ &\quad = -\frac{4\pi^2\hbar m}{(2\pi\hbar)^3}\int \mathrm{d}^3\mathbf{r}_1 \mathrm{d}^3\mathbf{r}_2\, \mathrm{e}^{-\mathrm{i}\mathbf{q}_2\cdot\mathbf{r}_2}\langle n', l', m'|\mathbf{r}_1\rangle\langle\mathbf{r}_1|n, l, m\rangle H_1(\mathbf{r}_1, \mathbf{r}_2), \end{aligned} \tag{12.119}$$

where $\mathbf{q}_2$ is the momentum transferred to the scattering electron. (Neglect the possibility that the two electrons exchange places. You may wish to perform the $\mathrm{d}^3\mathbf{r}_1$ integral by including a factor $\mathrm{e}^{-\alpha r_1}$ and then letting $\alpha \to 0$.)

Compute the differential cross-section for the $|1, 0, 0\rangle \to |2, 0, 0\rangle$ transition and show that at high energies it falls as $\mathrm{cosec}^{12}(\theta/2)$.

12.7 Use the optical theorem to show that the first Born approximation is not valid for forward scattering.

12.8 A particle scatters off a hard sphere, described by the potential

$$V(\mathbf{r}) = \begin{cases} \infty & \text{for } |\mathbf{r}| \le a \\ 0 & \text{otherwise.} \end{cases} \tag{12.120}$$

By considering the form of the radial wavefunction $u(r)$ in the region $r > a$, show that the phase shifts are given by $\tan\delta_l = j_l(ka)/n_l(ka)$, where $k = \sqrt{2mE}/\hbar$ and $j_l(kr)$ and $n_l(kr)$ are spherical Bessel functions and Neumann functions, which are the two independent solutions of the second-order radial equation

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}}{\mathrm{d}r}u(r)\right) = \left(\frac{l(l+1)}{r^2} - \frac{2mE}{\hbar^2}\right)u(r). \tag{12.121}$$

In the limit $kr \to 0$, show that these functions behave as

$$j_l(kr) \to \frac{(kr)^l}{2l+1} \qquad n_l(kr) \to -\frac{2l-1}{(kr)^{l+1}}\ . \tag{12.122}$$

Use this to show that in the low-energy limit, the scattering is spherically symmetric and the total cross-section is four times the classical value.

12.9 Show that in the Born approximation the phase shifts $\delta_l(E)$ for scattering off a spherical potential $V(r)$ are given by

$$\delta_l(E) \simeq -2mk\hbar^2\int_0^\infty \mathrm{d}r\, r^2 V(r)\,(j_l(kr))^2\,. \tag{12.123}$$

When is the approximation valid?

12.10 Two α-particles are collided. Show that when the α-particles initially have equal and opposite momenta, the differential cross-section is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = |f(\theta) + f(\theta - \pi)|^2. \tag{12.124}$$

Using the formula for $f(\theta)$ in terms of partial waves, show that the differential cross-section at $\theta = \pi/2$ is twice what would be expected had the α-particles been distinguishable.

A moving electron crashes into an electron that is initially at rest. Assuming both electrons are in the same spin state, show that the differential cross-section falls to zero at $\theta = \pi/4$.

Appendices

Appendix A: Cartesian tensors

Vector notation is very powerful, but sometimes it is necessary to step outside it and work explicitly with the components of vectors. This is especially true in quantum mechanics, because when operators are in play we have less flexibility about the order in which we write symbols, and standard vector notation can be prescriptive about order. For example if we want $\mathbf{p}$ to operate on $\mathbf{a}$ but not $\mathbf{b}$, we have to write $\mathbf{b}$ to the left of $\mathbf{p}$ and $\mathbf{a}$ on the right, but this requirement is incompatible with the vectorial requirements if the classical expression would be $\mathbf{p} \times (\mathbf{a} \times \mathbf{b})$. The techniques of Cartesian tensors resolve this kind of problem. Even in classical physics tensor notation enables us to use concepts that cannot be handled by vectors. In particular, it extends effortlessly to spaces with more than three dimensions, such as spacetime, which vector notation does only in a limited way.

Instead of writing $\mathbf{a}$, we write a_i for the ith component of the vector. Then $\mathbf{a}. \cdot \mathbf{b}$ becomes $\sum_i a_i b_i$. When a subscript is used twice in a product, as i is here, it is generally summed over and we speak of the subscript on a being **contracted** on the subscript on b.

The ijth component of the 3×3 identity matrix is denoted δ_{ij} and called the **Kronecker delta**: so

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

The equation $a_i = \sum_j \delta_{ij} a_j$ expresses the fact that the identity matrix times $\mathbf{a}$ equals $\mathbf{a}$. The scalar product often appears in the form $\mathbf{a} \cdot \mathbf{b} = \sum_{ij} \delta_{ij} a_i b_j$. To see that this is equivalent to the usual expression, we do the sum over j. Then the delta vanishes except when $j = i$, when it is unity, and we are left with $\sum_i q_i b_i$. Notice that it does not matter in what order the symbols appear; we have also $\mathbf{a} \cdot \mathbf{b} = \sum_{ij} a_i \delta_{ij} b_j$, etc. – when using Cartesian tensors, the information that in vector notation is encoded in the positions of symbols is carried by the way the subscripts on different symbols are coupled together.

To make the vector product we need to introduce the **alternating symbol** or **Levi–Civita symbol** ϵ_{ijk}. This is a set of 27 zeros and ones defined such that $\epsilon_{123} = 1$ and the sign changes if any two indices are swapped. So $\epsilon_{213} = -1$, $\epsilon_{231} = 1$, etc. If we cyclically permute the indices, changing 123 into first 312 and then 231, we are swapping two pairs each time, so there are two cancelling changes of sign. That is, $\epsilon_{123} = \epsilon_{312} = \epsilon_{231} = 1$ and $\epsilon_{213} = \epsilon_{321} = \epsilon_{132} = -1$. All the remaining 21 components of the alternating symbol vanish, because they have at least two subscripts equal, and swapping these equal subscripts we learn that this component is equal to minus itself, and therefore must be zero.

We now have

$$(\mathbf{a} \times \mathbf{b})_i = \sum_{jk} \epsilon_{ijk} a_j b_k. \tag{A.2}$$

To prove this statement, we explicitly evaluate the right side for $i = 1, 2$ and 3. For example, setting $i = 1$ the right side becomes $\sum_{jk} \epsilon_{1jk} a_j b_k$. In this sum ϵ_{1jk} is non-vanishing only when j is different from k and neither is equal one. So there are only two terms:

$$\epsilon_{123} a_2 b_3 + \epsilon_{132} a_3 a_2 = a_2 b_3 - a_3 b_2 \tag{A.3}$$

which is by definition the third component of $\mathbf{a} \times \mathbf{b}$.

A few simple rules enable us to translate between vector and tensor notation.

1. Fundamentally we are writing down the general component of some quantity, so if that quantity is a vector, there should be one subscript that is "spare" in the sense that it is not contracted with another subscript. Similarly, if the quantity is a scalar, all indices should be contracted, while a tensor quantity has two spare indices.
2. Each scalar product is expressed by choosing a letter that has not already been used for a subscript and making it the subscript of both the vectors involved in the product.
3. Each vector product is expressed by choosing three letters, say i, j and k and using them as subscripts of an ϵ. The second letter becomes the subscript that comes before the cross, and the third letter becomes the subscript of the vector that comes after the cross.

We need a lemma to handle vector triple products:

$$\sum_i \epsilon_{ijk} \epsilon_{irs} = \delta_{jr} \delta_{ks} - \delta_{kr} \delta_{js} \tag{A.4}$$

Before we prove this identity (which should be memorised), notice its shape: on the left we have two epsilons with a contracted subscript. On the right we have two products of deltas, the subscripts of which are matched "middle to middle, end to end" and "middle to end, end to middle". Now the proof. For the sum on the left to be non-vanishing, both epsilons must be non-vanishing for some value of i. For that value of i, the subscripts j and k must take the values that i does not. For example, if i is 1, j and k must between them be 2 and 3. For the same reason r and s must also between them be 2 and 3. So either $j = r$ and $k = s$ or $j = s$ and $k = r$. In the first case, if ijk is an even permutation of 123, then so is irs, or if ijk is an odd permutation, then so is irs. Hence in the first case either both epsilons are equal to 1, or they are both equal to -1 and their product is guaranteed to be 1. The first pair of deltas on the right cover this case. If, on the other hand, $j = s$ and $k = r$, irs will be an odd permutation of 123 if ijk is an even one, and vice versa if ijk is an odd permutation. Hence in this case one epsilon is equal to 1 and the other is equal to -1 and their product is certainly equal to -1. The second product of deltas covers this case. This completes the proof of equation (A.4) because we have shown that the two sides always take the same value no matter what values we give to the subscripts.

Besides enabling us to translate vector products into tensor notation, the alternating symbol enables us to form the determinant of any 3×3 matrix. In fact, this is the symbol's core role and its use for vector products is a spinoff from it. The simplest expression for $\det(\mathbf{A})$ is

$$\det(\mathbf{A}) = \sum_{ijk} \epsilon_{ijk} A_{1i} A_{2j} A_{3k}. \tag{A.5}$$

A more sophisticated expression that is often invaluable is

$$\det(\mathbf{A}) \epsilon_{rst} = \sum_{ijk} \epsilon_{ijk} A_{ri} A_{sj} A_{tk}. \tag{A.6}$$

These expressions are best treated as the definition of $\det(\mathbf{A})$ and used to derive the usual rule for the evaluation of a determinant by expansion down a row or column. This popular rule is actually a poor way to define a determinant, and a dreadful way of evaluating one. It should be avoided whenever possible.

Appendix B: Fourier series and transforms

The amplitude for a particle to be located at $\mathbf{x}$ and the amplitude for the particle to have momentum $\mathbf{p}$ are related by Fourier transforms, so they play a significant role in quantum mechanics. In this appendix we derive the basic formulae. Like Fourier[1] himself we start by considering a function of one variable, $f(x)$, that is periodic with period L: that is, $f(x+L) = f(x)$ for all x. We assume that f can be expressed as a sum of sinusoidal waves with wavelength L:

$$f(x) = \sum_{n=-\infty}^{\infty} F_n \mathrm{e}^{2\pi \mathrm{i} n x/L}, \tag{B.1}$$

where the F_n are complex numbers to be determined. At this stage this is just a hypothesis – only 127 years after Fourier introduced his series did Stone[2] prove that the sum on the right always converges to the function on the left. However numerical experiments soon convince us that the hypothesis is valid because it is straightforward to determine what the coefficients F_n must be, so we can evaluate them for some specimen functions f and see whether the series converges to the function. To determine the F_n we simply multiply both sides of the equation by $\mathrm{e}^{-2\pi \mathrm{i} m x/L}$ and integrate from $-L/2$ to $L/2$:[3]

$$\begin{aligned}\int_{-L/2}^{L/2} \mathrm{d}x\, \mathrm{e}^{-2\pi \mathrm{i} m x/L} f(x) &= \sum_{n=-\infty}^{\infty} F_n \int_{-L/2}^{L/2} \mathrm{d}x\, \mathrm{e}^{2\pi \mathrm{i}(n-m)x/L} \\ &= L F_m,\end{aligned} \tag{B.2}$$

where the second equality follows because for $n \neq m$ the integral of the exponential on the right vanishes, so there is only one non-zero term in the series. Thus the expansion coefficients have to be

$$F_m = \frac{1}{L}\int_{-L/2}^{L/2} \mathrm{d}x\, \mathrm{e}^{-2\pi \mathrm{i} m x/L} f(x). \tag{B.3}$$

In terms of the wavenumbers of our waves,

$$k_n \equiv \frac{2\pi n}{L}, \tag{B.4}$$

our formulae become

$$f(x) = \sum_{n=-\infty}^{\infty} F_n \mathrm{e}^{\mathrm{i} k_n x} \quad ; \quad F_m = \frac{1}{L}\int_{-L/2}^{L/2} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i} k_m x} f(x). \tag{B.5}$$

At this stage it proves expedient to replace the F_n with rescaled coefficients

$$\widetilde{f}(k_n) \equiv L F_n. \tag{B.6}$$

so our equations become

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{L}\widetilde{f}(k_n)\, \mathrm{e}^{\mathrm{i} k_n x} \quad ; \quad \widetilde{f}(k_m) = \int_{-L/2}^{L/2} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i} k_m x} f(x). \tag{B.7}$$

Now we eliminate L from the first equation in favour of the difference $\mathrm{d}k \equiv k_{n+1} - k_n = 2\pi/L$ and have

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi}\widetilde{f}(k_n)\, \mathrm{e}^{\mathrm{i} k_n x} \quad ; \quad \widetilde{f}(k_m) = \int_{-L/2}^{L/2} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i} k_m x} f(x). \tag{B.8}$$

Finally we imagine the period getting longer and longer without limit. As L grows the difference $\mathrm{d}k$ between successive values of k_n becomes smaller and smaller, so k_n becomes a continuous variable k, and the sum in the first equation of (B.8) becomes an integral. Hence in the limit of infinite L we are left with

$$f(x) = \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi}\, \widetilde{f}(k)\, \mathrm{e}^{\mathrm{i} k x} \quad ; \quad \widetilde{f}(k) = \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i} k x} f(x). \tag{B.9}$$

---

[1] After dropping out from a seminary Joseph Fourier (1768–1830) joined the Auxerre Revolutionary Committee. The Revolution's fratricidal violence led to his arrest but he avoided the guillotine by virtue of Robespierre's fall in 1794. He invented Fourier series while serving Napoleon as Prefect of Grenoble. His former teachers Laplace and Lagrange were not convinced.

[2] Marshall Stone strengthened a theorem proved by Karl Weierstrass in 1885.

[3] You can check that the integration can be over any interval of length L. We have chosen the interval $(-\frac{1}{2}L, \frac{1}{2}L)$ for later convenience.

These are the basic formulae of Fourier transforms. The original restriction to periodic functions has been lifted because any function can be considered to repeat itself after an infinite interval. The only restriction on f for these formulae to be valid is that it vanishes sufficiently fast at infinity for the integral in the second of equations (B.9) to converge: the requirement proves to be that $\int_{-\infty}^{\infty} \mathrm{d}x\, |f|^2$ exists, which requires that asymptotically $|f| < |x|^{-1/2}$. Physicists generally don't worry too much about this restriction.

Using the second of equations (B.9) to eliminate the Fourier transform $\widetilde{f}$ from the first equation, we have

$$f(x) = \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}x' \, \mathrm{e}^{\mathrm{i}k(x-x')} f(x'). \tag{B.10}$$

Mathematicians stop here because our next step is illegal.[4] Physicists reverse the order of the integrations in equation (B.10) and write

$$f(x) = \int_{-\infty}^{\infty} \mathrm{d}x' \, f(x') \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi} \, \mathrm{e}^{\mathrm{i}k(x-x')}. \tag{B.11}$$

Comparing this equation with equation (2.41) we see that the inner integral on the right satisfies the defining condition of the Dirac delta function, and we have

$$\delta(x - x') = \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{2\pi} \, \mathrm{e}^{\mathrm{i}k(x-x')}. \tag{B.12}$$

Appendix C: Operators in classical statistical mechanics

In classical statistical mechanics we are interested in the dynamics of a system with N degrees of freedom. We do not know the system's state, which would be quantified by its position $\mathbf{q} = (q_1, \dots, q_N)$ and the canonically conjugate momentum $\mathbf{p}$. Our limited knowledge is contained in the probability density $\psi(\mathbf{q}, \mathbf{p})$, which is defined such that the probability of the system being in the elementary phase-space volume $\mathrm{d}\tau = \mathrm{d}^N\mathbf{q}\, \mathrm{d}^N\mathbf{p}$ is $\psi(\mathbf{q}, \mathbf{p})\, \mathrm{d}\tau$.

Over time $\mathbf{q}$ and $\mathbf{p}$ evolve according to Hamilton's equations

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}} \quad ; \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}, \tag{C.1}$$

and ψ evolves such that probability is conserved:

$$\begin{aligned} 0 &= \frac{\partial \psi}{\partial t} + \frac{\partial}{\partial \mathbf{q}} \cdot (\dot{\mathbf{q}}\psi) + \frac{\partial}{\partial \mathbf{p}} \cdot (\dot{\mathbf{p}}\psi) \\ &= \frac{\partial \psi}{\partial t} + \frac{\partial H}{\partial \mathbf{p}} \cdot \frac{\partial \psi}{\partial \mathbf{q}} - \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{\partial \psi}{\partial \mathbf{p}} \\ &= \frac{\partial \psi}{\partial t} + \{\psi, H\}, \end{aligned} \tag{C.2}$$

where the second equality follows from substituting for $\dot{\mathbf{q}}$ and $\dot{\mathbf{p}}$ from Hamilton's equations, and the last line follows from the definition of a Poisson bracket: if $F(\mathbf{q}, \mathbf{p})$ and $G(\mathbf{q}, \mathbf{p})$ are any two functions on phase space, the **Poisson bracket** $\{F, G\}$ is defined to be

$$\{F, G\} \equiv \sum_i \frac{\partial F}{\partial q_i} \frac{\partial G}{\partial p_i} - \frac{\partial F}{\partial p_i} \frac{\partial G}{\partial q_i}. \tag{C.3}$$

We use the Poisson bracket to associate with F an operator $\hat{F}$ on other functions of phase space: we define the operator $\hat{F}$ by its action on an arbitrary function $\psi(\mathbf{p}, \mathbf{q})$:

$$\hat{F}\psi \equiv -\mathrm{i}\hbar \{\psi, F\}. \tag{C.4}$$

Here $\hbar$ is some constant with the dimensions of the product $\mathbf{p} \cdot \mathbf{q}$ – i.e., the inverse of the dimensions of the Poisson bracket – and is introduced so the operator $\hat{F}$ has

---

[4] It is legitimate to reverse the order of integration only when the integrand is absolutely convergent, i.e., the integral of its absolute value is finite. This condition is clearly not satisfied in the case of $\mathrm{e}^{\mathrm{i}kx}$. By breaking the proper rules we obtain an expression for an object, the Dirac delta function, that is not a legitimate function. However, it *is* extremely useful.

Box C.1: Classical operators for a single particle

In the simplest case, our system consists of a single particle with Hamiltonian $H = \frac{1}{2}p^2/m + V(\mathbf{x})$. Then the operators associated with p_x, x, H and L_z are

$$\begin{aligned}\hat{p}_x &= -\mathrm{i}\hbar\{\cdot, p_x\} = -\mathrm{i}\hbar\frac{\partial}{\partial x} \quad ; \quad \hat{x} = -\mathrm{i}\hbar\{\cdot, x\} = \mathrm{i}\hbar\frac{\partial}{\partial p_x}\\ \hat{H} &= -\mathrm{i}\hbar\{\cdot, H\} = -\mathrm{i}\hbar\Big(\frac{\mathbf{p}}{m}\cdot\nabla - \nabla V\cdot\frac{\partial}{\partial\mathbf{p}}\Big)\\ \hat{L}_z &= -\mathrm{i}\hbar\{\cdot, L_z\} = -\mathrm{i}\hbar\{\cdot, xp_y - yp_x\}\\ &= -\mathrm{i}\hbar\Big(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} + p_x\frac{\partial}{\partial p_y} - p_y\frac{\partial}{\partial p_x}\Big).\end{aligned} \tag{1}$$

Notice that $\widehat{(p^2)} \neq (\hat{p})^2$. The commutators of these operators are

$$[\hat{x}, \hat{p}_x] = 0 \quad ; \quad [\hat{L}_x, \hat{L}_y] = \mathrm{i}\hbar\hat{L}_z. \tag{2}$$

(The easiest way to derive the second of these results is to apply (C.8).)

the same dimensions as the function F. The factor i is introduced to ensure that with the obvious definition of an inner product, the operator $\hat{F}$ is Hermitian:

$$\begin{aligned}\int \mathrm{d}\tau\,\phi^*\hat{F}\psi &= -\mathrm{i}\hbar\Big(\int \mathrm{d}^N\mathbf{p}\,\mathrm{d}^N\mathbf{q}\,\phi^*\frac{\partial\psi}{\partial\mathbf{q}}\cdot\frac{\partial F}{\partial\mathbf{p}} - \int \mathrm{d}^N\mathbf{q}\,\mathrm{d}^N\mathbf{p}\,\phi^*\frac{\partial\psi}{\partial\mathbf{p}}\cdot\frac{\partial F}{\partial\mathbf{q}}\Big)\\ &= \mathrm{i}\hbar\Big(\int \mathrm{d}^N\mathbf{p}\,\mathrm{d}^N\mathbf{q}\,\frac{\partial\phi^*}{\partial\mathbf{q}}\cdot\frac{\partial F}{\partial\mathbf{p}}\psi - \int \mathrm{d}^N\mathbf{q}\,\mathrm{d}^N\mathbf{p}\,\frac{\partial\phi^*}{\partial\mathbf{p}}\cdot\frac{\partial F}{\partial\mathbf{q}}\psi\Big)\\ &= \int \mathrm{d}\tau\,(\hat{F}\phi)^*\psi.\end{aligned} \tag{C.5}$$

When written in terms of the classical Hamiltonian operator $\hat{H}$, the classical evolution equation (C.2) takes a familiar form

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi. \tag{C.6}$$

It is straightforward (if tedious) to show that Poisson brackets, like a commutators, satisfies the Jacobi identity (cf. 2.92)

$$\{A, \{B, C\}\} + \{B, \{C, A\}\} + \{C, \{A, B\}\} = 0. \tag{C.7}$$

We use this identity to express the commutator of two such operators in terms of the Poisson bracket of the underlying functions:

$$\begin{aligned}[\hat{A}, \hat{B}]\psi &= -\hbar^2\big(\{\{\psi, B\}, A\} - \{\{\psi, A\}, B\}\big)\\ &= \hbar^2\{\psi, \{A, B\}\}\\ &= \mathrm{i}\hbar\widehat{\{A, B\}}\psi.\end{aligned} \tag{C.8}$$

where $\widehat{\{A, B\}}$ denotes the operator associated with the function $\{A, B\}$.

Let $A(\mathbf{p}, \mathbf{q})$ be some function on phase space. Then the rate of change of the value of A along a phase trajectory is

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{\partial A}{\partial\mathbf{p}}\cdot\dot{\mathbf{p}} + \frac{\partial A}{\partial\mathbf{q}}\cdot\dot{\mathbf{q}} = \{A, H\}. \tag{C.9}$$

Consequently A is a constant of motion if and only if $0 = \{A, H\}$, which by (C.8) requires its operator to commute with the Hamiltonian operator: as in quantum mechanics, A is a constant of the motion if and only if $[\hat{A}, \hat{H}] = 0$.

It is instructive to repeat the analysis of §4.2 with the classical operators of a single-particle system (Box C.1). If we displace the system through $\mathbf{a}$, its probability density becomes

$$\begin{aligned}\psi'(\mathbf{x}, \mathbf{p}) \equiv \psi(\mathbf{x} - \mathbf{a}, \mathbf{p}) &= \left[1 - \mathbf{a}\cdot\frac{\partial}{\partial\mathbf{x}} + \frac{1}{2!}\left(\mathbf{a}\cdot\frac{\partial}{\partial\mathbf{x}}\right)^2 - \dots\right]\psi(\mathbf{x}, \mathbf{p})\\ &= \exp\left(-\mathbf{a}\cdot\frac{\partial}{\partial\mathbf{x}}\right)\psi(\mathbf{x}, \mathbf{p}) = \exp\left(-\mathrm{i}\frac{\mathbf{a}\cdot\hat{\mathbf{p}}}{\hbar}\right)\psi(\mathbf{x}, \mathbf{p}).\end{aligned} \tag{C.10}$$

Thus $\hat{\mathbf{p}}/\hbar$ is the generator of displacements, as in quantum mechanics. Displacement of the system by $\delta\mathbf{a}$ clearly increases the expectation of $\mathbf{x}$ by $\delta\mathbf{a}$, so with

$\mathrm{d}\tau \equiv \mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{p}$

$$\langle\mathbf{x}\rangle + \delta\mathbf{a} = \int \mathrm{d}\tau\,\mathbf{x}\psi'(\mathbf{x},\mathbf{p}) = \int \mathrm{d}\tau\,\mathbf{x}\left(1 - \mathrm{i}\frac{\delta\mathbf{a}\cdot\hat{\mathbf{p}}}{\hbar}\right)\psi(\mathbf{x},\mathbf{p}) + \mathrm{O}(\delta a^2). \tag{C.11}$$

This equation will hold for an arbitrary probability density ψ if and only if

$$\mathrm{i}\hbar\delta_{ij} = \int \mathrm{d}\tau\, x_i\hat{p}_j\psi = \int \mathrm{d}\tau\,(\hat{p}_j x_i)^*\psi = \mathrm{i}\hbar\int \mathrm{d}\tau\,\{x_i, p_j\}\psi, \tag{C.12}$$

where the second equality uses the fact that $\hat{p}_j$ is Hermitian. Thus equation (C.11) holds if and and only if the Poisson brackets $\{x_i, p_j\}$ rather than the commutators $[\hat{x}_i, \hat{p}_j]$ satisfy the canonical commutation relations. This crucial difference between the quantum and classical cases arises from the way we calculate expectation values: in classical physics the quantum rule $\langle Q\rangle = \langle\psi|Q|\psi\rangle$ is replaced by

$$\langle Q\rangle = \int \mathrm{d}^N\mathbf{q}\,\mathrm{d}^N\mathbf{p}\,Q\psi, \tag{C.13}$$

where (i) Q is the function not the associated operator, and (ii) ψ occurs once not twice because it is a probability not a probability amplitude. On account of these differences, whereas equation (4.20) yields $[x_i, p_j] = \mathrm{i}\hbar\delta_{ij}$, its classical analogue, (C.11) yields $\{x_i, p_j\} = \delta_{ij}$.

Appendix D: Lorentz covariant equations

Special relativity is about determining how the numbers used by moving observers to characterise a given system are related to one another. All observers agree on some numbers such as the electric charge on a particle – these numbers are called **Lorentz scalars**. Other numbers, such as energy, belong to a set of four numbers, called a **four-vector**. If you know the values assigned by some observer to all four components of a four-vector, you can predict the values that any other observer will assign. If you do not know *all* four numbers, in general you cannot predict *any* of the values that a moving observer will assign. The components of every ordinary three-dimensional vector are associated with the spatial components of a four-vector, while some other number completes the set as the 'time component' of the four-vector. We use the convention that Greek indices run from 0 to 3 while Roman ones run from 1 to 3; the time component is component 0, followed by the x component, and so on. All components should have the same dimensions, so, for example, the energy-momentum four vector is

$$(p^0, p^1, p^2, p^3) \equiv (E/c, p_x, p_y, p_z). \tag{D.1}$$

The energies and momenta assigned by an observer who moves at speed v parallel to the x axis are found by multiplying the four-vector by a Lorentz transformation matrix. For example, if the primed observer is moving at speed v along the x axis, then she measures

$$\begin{pmatrix} E'/c \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E/c \\ p_x \\ p_y \\ p_z \end{pmatrix}, \tag{D.2}$$

where $\beta \equiv v/c$ and the **Lorentz factor** is $\gamma = 1/\sqrt{1-\beta^2}$. The indices on the four-vector $\mathbf{p}$ are written as superscripts because it proves helpful to have a form of $\mathbf{p}$ in which the sign of the time component is reversed. That is we define

$$(p_0, p_1, p_2, p_3) \equiv (-E/c, p_x, p_y, p_z), \tag{D.3}$$

and we write the indices on the left as subscripts to signal the difference in the time component. It is straightforward to verify that in the primed frame the components of the down vector are obtained by multiplication with a matrix that differs slightly from the one used to transform the up vector

$$\begin{pmatrix} -E'/c \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -E/c \\ p_x \\ p_y \\ p_z \end{pmatrix}. \tag{D.4}$$

The Lorentz transformation matrices that appear in equation (D.2) and (D.4) are inverses of one another. In index notation we write these equations

$$p'^{\nu} = \sum_\mu \Lambda^\nu{}_\mu p^\mu \quad \text{and} \quad p'_\nu = \sum_\mu \Lambda_\nu{}^\mu p_\mu. \tag{D.5}$$

Notice that we sum over one down and one up index; we never sum over two down or two up indices. Summing over an up and down index is called **contraction** of those indices.

The dot product of the up and down forms of a four vector yields a Lorentz scalar. For example

$$\sum_\mu p_\mu p^\mu = -\frac{E^2}{c^2} + p_x^2 + p_y^2 + p_z^2 = -m_0^2 c^2, \tag{D.6}$$

where m_0 is the particle's rest mass. The dot product of two different four vectors is also a Lorentz scalar: the value of $\sum_\mu p_\mu h^\mu = \sum_\mu p^\mu h_\mu$ is the same in any frame.

The time and space coordinates form a four-vector

$$(x^0, x^1, x^2, x^3) \equiv (ct, x, y, z). \tag{D.7}$$

In some interval $\mathrm{d}t$ of coordinate time, a particle's position four-vector $\mathbf{x}$ increments by $\mathrm{d}\mathbf{x}$ and the Lorentz scalar associated with $\mathrm{d}\mathbf{x}$ is $-c^2$ times the square of the **proper-time interval** associated with $\mathrm{d}t$:

$$\begin{aligned}(\mathrm{d}\tau)^2 = -\frac{1}{c^2}\sum_\mu \mathrm{d}x_\mu \mathrm{d}x^\mu &= (\mathrm{d}t)^2 - \frac{1}{c^2}\left\{(\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2\right\} \\ &= (\mathrm{d}t)^2(1-\beta^2).\end{aligned} \tag{D.8}$$

The proper timeime: proper $\mathrm{d}\tau$ is just the elapse of time in the particle's instantaneous rest frame; it is the amount by which the hands move on a clock that is tied to the particle. From the last equality above it follows that $\mathrm{d}\tau = \mathrm{d}t/\gamma$, so moving clocks tick slowly.

The **four-velocity** of a particle is

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \left(\frac{\mathrm{d}ct}{\mathrm{d}\tau}, \frac{\mathrm{d}x}{\mathrm{d}\tau}, \frac{\mathrm{d}y}{\mathrm{d}\tau}, \frac{\mathrm{d}z}{\mathrm{d}\tau}\right) = \gamma\left(c, \frac{\mathrm{d}x}{\mathrm{d}t}, \frac{\mathrm{d}y}{\mathrm{d}t}, \frac{\mathrm{d}z}{\mathrm{d}t}\right), \tag{D.9}$$

where γ is the particle's Lorentz factor. In a particle's rest frame the four velocity points straight into the future: $u^\mu = (1, 0, 0, 0)$. In any frame

$$u_\mu u^\mu = -c^2. \tag{D.10}$$

Some numbers are members of a set of six numbers that must all be known in one frame before any of them can be predicted in an arbitrary frame. The six components of the electric and magnetic fields form such a set. We arrange them as the independent, non-vanishing components of an antisymmetric four by four matrix, called the **Maxwell field tensor**

$$F_{\mu\nu} \equiv \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}. \tag{D.11}$$

The electric and magnetic fields seen by a moving observer are obtained by pre- and post-multiplying this matrix by an appropriate Lorentz transformation matrix such as that appearing in equation (D.4).

The equation of motion of a particle of rest mass m_0 and charge Q is

$$m_0 \frac{\mathrm{d}u_\lambda}{\mathrm{d}\tau} = Q \sum_\nu F_{\lambda\nu} u^\nu. \tag{D.12}$$

The time component of the four-velocity $\mathbf{u}$ is γc, and the spatial part is $\gamma \mathbf{v}$, so, using our expression (D.11) for $\mathbf{F}$, the spatial part of this equation of motion is

$$\gamma Q(\mathbf{E} + \mathbf{v}\times\mathbf{B}) = m_0 \frac{\mathrm{d}\gamma\mathbf{v}}{\mathrm{d}\tau} = \gamma m_0 \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\tau} + \mathrm{O}(\beta^2), \tag{D.13}$$

which shows the familiar electrostatic and Lorentz forces in action.

The great merit of establishing these rules is that we can state that the dynamics of any system can be determined from equations in which both sides are of the same Lorentz-covariant type. That is, both sides are Lorentz scalars, or four-vectors, or antisymmetric matrices, or whatever. Any correct equation that does not conform to this pattern must be a fragment of a set of equations that do. Once a system's governing equations have been written in Lorentz covariant form, we can instantly transform them to whatever reference frame we prefer to work in.

Appendix E: **Thomas precession**

In this appendix we generalise the equation of motion of an electron's spin (eq. 8.65)

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{gQ}{2m_0}\mathbf{S}\times\mathbf{B} \tag{E.1}$$

from the electron's rest frame to a frame in which the electron is moving. We do this by writing equation (E.1) in Lorentz covariant form (Appendix D).

The first step in upgrading equation (E.1) to Lorentz covariant form is to replace $\mathbf{S}$ and $\mathbf{B}$ with covariant structures. We hypothesise that the numbers S_i comprise the spatial components of a four vector that has vanishing time component in the particle's rest frame. Thus

$$(s^0, s^1, s^2, s^3) = (0, S_x, S_y, S_z) \quad \text{(rest frame)}, \tag{E.2}$$

and we can calculate s^μ in an arbitrary frame by multiplying this equation by an appropriate Lorentz transformation matrix. Since in the rest frame s^μ is orthogonal to the particle's four-velocity u^μ, the equation

$$u_\mu s^\mu = 0 \tag{E.3}$$

holds in any frame. In equation (8.65) $\mathbf{B}$ clearly has to be treated as part of the Maxwell field tensor $F_{\mu\nu}$ (eq. D.11). In the particle's rest frame $\mathrm{d}t = \mathrm{d}\tau$ and

$$\sum_\nu F_{\mu\nu}s^\nu = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}\begin{pmatrix} 0 \\ S_x \\ S_y \\ S_z \end{pmatrix} = \begin{pmatrix} -\mathbf{S}\cdot\mathbf{E}/c \\ \\ \mathbf{S}\times\mathbf{B} \\ \\ \end{pmatrix} \tag{E.4}$$

so equation (E.1) coincides with the spatial components of the covariant equation

$$\frac{\mathrm{d}s_\mu}{\mathrm{d}\tau} = \frac{gQ}{2m_0}\sum_\nu F_{\mu\nu}s^\nu. \tag{E.5}$$

This cannot be the correct equation, however, because it is liable to violate the condition (E.3). To see this, consider the case in which the particle moves at constant velocity and dot equation (E.5) through by the fixed four-velocity u^ν. Then we obtain

$$\sum_\mu \frac{\mathrm{d}}{\mathrm{d}\tau}(u^\mu s_\mu) = \frac{gQ}{2m_0}\sum_{\mu\nu} F_{\mu\nu}u^\mu s^\nu. \tag{E.6}$$

The left side has to be zero but there is no reason why the right side should vanish. We can fix this problem by adding an extra term to the right side, so that

$$\frac{\mathrm{d}s_\mu}{\mathrm{d}\tau} = \frac{gQ}{2m_0}\left(\sum_\nu F_{\mu\nu}s^\nu - \sum_{\lambda\nu} s^\lambda F_{\lambda\nu}\frac{u^\nu u_\mu}{c^2}\right). \tag{E.7}$$

When this equation is dotted through by u^μ, and equation (D.10) is used, the right side becomes proportional to $\sum_{\mu\nu} F_{\mu\nu}(s^\mu u^\nu + s^\nu u^\mu)$, which vanishes because $\mathbf{F}$ is antisymmetric in its indices while the bracket into which it is contracted is symmetric in the same indices.[1]

If our particle is accelerating, equation (E.7) is still incompatible with equation (E.3), as becomes obvious when one dots through by u^μ and includes a non-zero term $\mathrm{d}u^\mu/\mathrm{d}\tau$. Fortunately, this objection is easily fixed by adding a third term to the right side. We then have our final covariant equation of motion for $\mathbf{s}$

$$\frac{\mathrm{d}s_\mu}{\mathrm{d}\tau} = \frac{gQ}{2m_0}\left(\sum_\nu F_{\mu\nu}s^\nu - \sum_{\lambda\nu} s^\lambda F_{\lambda\nu}\frac{u^\nu u_\mu}{c^2}\right) + \frac{1}{c^2}\sum_\lambda s^\lambda \frac{\mathrm{d}u_\lambda}{\mathrm{d}\tau}u_\mu. \tag{E.8}$$

In the rest frame the spatial components of this covariant equation coincide with the equation (8.65) that we started from because $u_i = 0$. The two new terms on the right side ensure that $\mathbf{s}$ remains perpendicular to the four-velocity $\mathbf{u}$ as it must do if it is to have vanishing time component in the rest frame.

The last term on the right of equation (E.8) is entirely generated by the particle's acceleration; it would survive even in the case $g = 0$ of vanishing magnetic

---

[1] Here's a proof that the contraction of tensors $\mathbf{S}$ and $\mathbf{A}$ that are respectively symmetric and antisymmetric in their indices vanishes. $\sum_{\mu\nu} S_{\mu\nu}A_{\mu\nu} = \sum_{\mu\nu} S_{\nu\mu}A_{\mu\nu} = -\sum_{\mu\nu} S_{\nu\mu}A_{\nu\mu}$. This establishes that the sum is equal to minus itself. Zero is the only number that has this property.

moment. Thus the spin of an accelerating particle precesses regardless of torques. This precession is called **Thomas precession**.[2]

If the particle's acceleration is entirely due to the electromagnetic force that it experiences because it is charged, its equation of motion is (D.12). Using this in equation (E.8), we find

$$\frac{\mathrm{d}s_\mu}{\mathrm{d}\tau} = \frac{Q}{2m_0}\left(g\sum_\nu F_{\mu\nu}s^\nu - (g-2)\sum_{\lambda\nu} s^\lambda F_{\lambda\nu}\frac{u^\nu u_\mu}{c^2}\right). \tag{E.9}$$

For electrons, $g = 2.002$ and to a good approximation the extra terms we have added cancel and our originally conjectured equation (E.5) holds after all. We now specialise on the unusually simple and important case in which $g = 2$.

From our equation of motion of the covariant object $\mathbf{s}$ we derive the equation of motion of the three-vector $\mathbf{S}$ whose components are the expectation values of the spin operators. We choose to work in the rest frame of the atom. By equation (E.2), $\mathbf{S}$ is related to $\mathbf{s}$ by a time-dependent Lorentz transformation from this frame to the electron's rest frame. We align our x axis with the direction of the required boost, so

$$\begin{pmatrix} 0 \\ S_x \\ S_y \\ S_z \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s^0 \\ s^1 \\ s^2 \\ s^3 \end{pmatrix}. \tag{E.10}$$

The time equation implies that $s^0 = \beta s^1$, so the x equation can be written

$$S_x = \gamma(s^1 - \beta s^0) = \gamma(1-\beta^2)s^1 = \frac{s^1}{\gamma} = s^1(1 - \tfrac{1}{2}\beta^2 + \cdots). \tag{E.11}$$

The y and z components of equation (E.10) state that the corresponding components of $\mathbf{S}$ and $\mathbf{s}$ are identical. Since s^1 is the projection of the spatial part of $\mathbf{s}$ onto the particle's velocity $\mathbf{v}$, we can summarise these results in the single equation

$$\mathbf{S} = \mathbf{s} - \frac{\mathbf{v}\cdot\mathbf{s}}{2c^2}\mathbf{v} + \mathrm{O}(\beta^4) \tag{E.12}$$

as one can check by dotting through with the unit vectors in the three spatial directions. Differentiating with respect to proper time and discarding terms of order β^2, we find

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}\tau} = \frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\tau} - \frac{1}{2c^2}\left(\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\tau}\cdot\mathbf{s}\right)\mathbf{v} - \frac{\mathbf{v}\cdot\mathbf{s}}{2c^2}\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}\tau} + \mathrm{O}(\beta^2). \tag{E.13}$$

Equation (E.9) implies that with $g = 2$

$$\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\tau} = \frac{Q}{m_0}\left(\frac{s^0}{c}\mathbf{E} + \mathbf{s}\times\mathbf{B}\right) = \frac{Q}{m_0}\left(\frac{\mathbf{v}\cdot\mathbf{s}}{c^2}\mathbf{E} + \mathbf{s}\times\mathbf{B}\right), \tag{E.14}$$

where the second equality uses the relation $s^0 = \beta s^1 = (\mathbf{v}\cdot\mathbf{s})/c$. We now use this equation and equation (D.13) to eliminate $\mathrm{d}\mathbf{s}/\mathrm{d}\tau$ and $\mathrm{d}\mathbf{v}/\mathrm{d}\tau$ from equation (E.13).

$$\begin{aligned} \frac{\mathrm{d}\mathbf{S}}{\mathrm{d}\tau} &= \frac{Q}{2m_0c^2}\big((\mathbf{v}\cdot\mathbf{s})\mathbf{E} - (\mathbf{E}\cdot\mathbf{s})\mathbf{v} + 2c^2\mathbf{s}\times\mathbf{B}\big) + \mathrm{O}(\beta^2) \\ &= \frac{Q}{2m_0c^2}\big(\mathbf{S}\times(\mathbf{E}\times\mathbf{v}) + 2c^2\mathbf{S}\times\mathbf{B}\big) + \mathrm{O}(\beta^2). \end{aligned} \tag{E.15}$$

Since we are working in the atom's rest frame, $\mathbf{B} = 0$ unless we are applying an external electric field. The difference between the electron's proper time τ and the atom's proper time t is $\mathrm{O}(\beta^2)$, so we can replace τ with t. We assume that $\mathbf{E}$ is generated by an electrostatic potential $\Phi(r)$ that is a function of radius only. Then $\mathbf{E} = -\nabla\Phi = -(\mathrm{d}\Phi/\mathrm{d}r)\mathbf{r}/r$ points in the radial direction. Using this relation in equation (E.15) we find

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{Q}{2m_0c^2}\left(-\frac{1}{r}\frac{\mathrm{d}\Phi}{\mathrm{d}r}\mathbf{S}\times(\mathbf{r}\times\mathbf{v}) + 2c^2\mathbf{S}\times\mathbf{B}\right). \tag{E.16}$$

When $\mathbf{r}\times\mathbf{v}$ is replaced by $\hbar\mathbf{L}/m_0$, we obtain equation (8.66). The factor of two difference between the coefficients of $\mathbf{S}$ in the spin-orbit and Zeeman Hamiltonians (8.67) and (8.68) that so puzzled the pioneers of quantum mechanics, arises because the variable in equation (E.5) is not the usual spin operator but a Lorentz transformed version of it. The required factor of two emerges from the derivatives of $\mathbf{v}$ in equation (E.13). Hence it is a consequence of the fact that the electron's rest frame is accelerating.

---

[2] L.T. Thomas, *Phil. Mag.* **3**, 1 (1927). For an illuminating discussion see §11.11 of *Classical Electrodynamics* by J.D. Jackson (Wiley).

Appendix F: Matrix elements for a dipole-dipole interaction

We calculate the matrix elements obtained by squeezing the hyperfine Hamiltonian (8.77) between states that would all be ground states if the nucleus had no magnetic dipole moment. We assume that these are S states, and therefore have a spherically symmetric spatial wavefunction $\psi(r)$. They differ only in their spins. In practice they will be the eigenstates $|j,m\rangle$ of the total angular momentum operators J^2 and J_z that can be constructed by adding the nuclear and electron spins. We use the symbol s as a shorthand for j,m or whatever other eigenvalues we decide to use.

The matrix elements are

$$M_{ss'} \equiv \langle\psi, s|H_{\rm HFS}|\psi, s'\rangle = \int {\rm d}^3\mathbf{x}\,\rho(r)\langle s|H_{\rm HFS}|s'\rangle, \tag{F.1a}$$

where

$$\rho(r) \equiv |\psi(r)|^2, \tag{F.1b}$$

and for given s, s' $\langle s|H_{\rm HFS}|s'\rangle$ is a function of position $\mathbf{x}$ only. Substituting for $H_{\rm HFS}$ from equation (8.77) we have

$$M_{ss'} = \frac{\mu_0}{4\pi}\int {\rm d}^3\mathbf{x}\,\rho\langle s|\boldsymbol{\mu}_{\rm N}\cdot\nabla\times\left\{\nabla\times\left(\frac{\boldsymbol{\mu}_{\rm e}}{r}\right)\right\}|s'\rangle. \tag{F.2}$$

We now use tensor notation (Appendix A) to extract the spin operators from the integral, finding

$$M_{ss'} = \frac{\mu_0}{4\pi}\sum_{ijklm}\epsilon_{ijk}\epsilon_{klm}\langle s|\mu_{{\rm N}i}\mu_{{\rm e}m}|s'\rangle I, \tag{F.3a}$$

where

$$I \equiv \int {\rm d}^3\mathbf{x}\,\rho(r)\frac{\partial^2 r^{-1}}{\partial x_j\partial x_l}. \tag{F.3b}$$

The domain of integration is a large sphere centred on the origin. On evaluating the derivatives of r^{-1} and writing the volume element ${\rm d}^3\mathbf{x}$ in spherical polar coordinates, the integral becomes

$$I = \int \rho(r)r^2{\rm d}r\int {\rm d}^2\Omega\left(3\frac{x_jx_l}{r^5} - \frac{\delta_{jl}}{r^3}\right). \tag{F.4}$$

We integrate over polar angles first. If $j \neq l$, the first term integrates to zero because the contribution from a region in which x_j is positive is exactly cancelled by a contribution from a region in which x_j is negative. When $j = l$, we orient our axes so that x_j is the z axis. Then the angular integral becomes

$$\int {\rm d}\Omega\left(3\frac{x_jx_l}{r^5} - \frac{\delta_{jl}}{r^3}\right) = \frac{2\pi}{r^3}\int {\rm d}\theta\,\sin\theta(3\cos^2\theta - 1) = 0. \tag{F.5}$$

The vanishing of the angular integral implies that no contribution to the integral of equation (F.3b) comes from the entire region $r > 0$. However, we cannot conclude that the integral vanishes entirely because the coefficient of ρ in the radial integral of (F.4) is proportional to $1/r$, so the radial integral is divergent for $\rho(0) \neq 0$.

Since any contribution comes from the immediate vicinity of the origin, we return to our original expression but restrict the region of integration to an infinitesimal sphere around the origin. We approximate $\rho(r)$ by $\rho(0)$ and take it out of the integral. Then we can invoke the divergence theorem to state that

$$I = \rho(0)\int {\rm d}^3\mathbf{x}\,\frac{\partial^2 r^{-1}}{\partial x_j\partial x_l} = \int {\rm d}^2\Omega\,rx_j\frac{\partial r^{-1}}{\partial x_l}, \tag{F.6}$$

where we have used the fact that on the surface of a sphere of radius r the infinitesimal surface element is ${\rm d}^2\mathbf{S} = {\rm d}^2\Omega\,r\mathbf{x}$. We now evaluate the surviving derivative of $1/r$:

$$I = -\rho(0)\int {\rm d}^2\Omega\,\frac{x_jx_l}{r^2} = -\frac{4\pi}{3}\rho(0)\delta_{jl}, \tag{F.7}$$

where we have again exploited the fact that the integral vanishes by symmetry if $j \neq l$, and that when $j = l$ it can be evaluated by taking x_j to be z. Inserting this value of I in equation (F.3a), we have

$$M_{ss'} = -\frac{\mu_0}{3}\rho(0)\sum_{ijklm}\epsilon_{ijk}\epsilon_{klm}\delta_{jl}\langle s|\mu_{{\rm N}i}\mu_{{\rm e}m}|s'\rangle. \tag{F.8}$$

Now

$$\sum_{ijklm}\epsilon_{ijk}\epsilon_{klm}\delta_{jl} = \sum_{ijkm}\epsilon_{ijk}\epsilon_{kjm} = -\sum_{ijkm}\epsilon_{ijk}\epsilon_{mjk} \tag{F.9}$$

This sum must be proportional to δ_{im} because if $i \neq m$, it is impossible for both (ijk) and (mjk) to be permutations of (xyz) as they must be to get a non-vanishing contribution to the sum. We can determine the constant of proportionality by making a concrete choice for $i = m$. For example, when they are both x we have

$$\sum_{jk} \epsilon_{xjk}\epsilon_{xjk} = \epsilon_{xyz}\epsilon_{xyz} + \epsilon_{xzy}\epsilon_{xzy} = 2. \tag{F.10}$$

When these results are used in equation (F.8), we have finally

$$M_{ss'} = \frac{2\mu_0}{3}|\psi(0)|^2 \langle s| \sum_i \mu_{\mathrm{N}i} \cdot \mu_{\mathrm{e}i} |s'\rangle. \tag{F.11}$$

Appendix G: Selection rule for j

In Problem 7.24 the selection rule on l is derived by calculating $[L^2, [L^2, x_i]]$ and then squeezing the resulting equation between states $\langle l, m_l|$ and $|l', m_l'\rangle$. The algebra uses only two facts about the operators $\mathbf{L}$ and $\mathbf{x}$, namely $[L_i, x_j] = \mathrm{i}\sum_k \epsilon_{ijk}x_k$, and $\mathbf{L}\cdot\mathbf{x} = 0$. Now if we substitute $\mathbf{J}$ for $\mathbf{L}$, the first equation carries over (i.e., $[J_i, x_j] = \mathrm{i}\sum_k \epsilon_{ijk}x_k$) but the second does not ($\mathbf{J}\cdot\mathbf{x} = \mathbf{S}\cdot\mathbf{x}$). To proceed, we define the operator

$$\mathbf{X} \equiv \mathbf{J} \times \mathbf{x} - \mathrm{i}\mathbf{x}. \tag{G.1}$$

Since $\mathbf{X}$ is a vector operator, it will satisfy the commutation relations $[J_i, X_j] = \mathrm{i}\sum_k \epsilon_{ijk}X_k$, as can be verified by explicit calculation. Moreover $\mathbf{X}$ *is* perpendicular to $\mathbf{J}$:

$$\begin{aligned}\mathbf{J}\cdot\mathbf{X} &= \sum_{klm} \epsilon_{klm} J_k J_l x_m - \mathrm{i}\sum_m J_m x_m \\ &= \tfrac{1}{2}\sum_{klm} \epsilon_{klm} [J_k, J_l] x_m - \mathrm{i}\sum_m J_m x_m \\ &= \tfrac{1}{2}\sum_{klm} \epsilon_{klm} \mathrm{i}\sum_p \epsilon_{klp} J_p x_m - \mathrm{i}\sum_m J_m x_m = 0,\end{aligned} \tag{G.2}$$

where the last equality uses equation (F.10). We can now argue that the algebra of Problem 7.24 will carry over with $\mathbf{J}$ substituted for $\mathbf{L}$ and $\mathbf{X}$ substituted for $\mathbf{x}$. Hence the matrix elements $\langle jm|X_k|j'm'\rangle$ satisfy the selection rule $|j - j'| = 1$.

Now we squeeze a component of equation (G.1) between two states of well-defined angular momentum

$$\begin{aligned}\langle jm|X_r|j'm'\rangle &= \sum_{st} \epsilon_{rst} \sum_{j''m''} \langle jm|J_s|j''m''\rangle\langle j''m''|x_t|j'm'\rangle - \mathrm{i}\langle jm|x_r|j'm'\rangle \\ &= \sum_{m''st} \epsilon_{rst} \langle jm|J_s|jm''\rangle\langle jm''|x_t|j'm'\rangle - \mathrm{i}\langle jm|x_r|j'm'\rangle,\end{aligned} \tag{G.3}$$

where the sum over j'' has been reduced to the single term $j'' = j$ because $[J^2, J_s] = 0$. The left side vanishes for $|j - j'| \neq 1$. Moreover, since $\mathbf{J}\cdot\mathbf{x}$ is a scalar, it commutes with J^2 and we have that $\langle jm|\mathbf{J}\cdot\mathbf{x}|j'm'\rangle = 0$ unless $j = j'$, or

$$\sum_{m''t} \langle jm|J_t|jm''\rangle\langle jm''|x_t|j'm'\rangle \propto \delta_{jj'} \tag{G.4}$$

Let $|j - j'| > 1$, then in matrix notation we can write equations (G.3) for $r = x, y$ and (G.4) as

$$\begin{aligned}0 &= \mathbf{J}_y\mathbf{z} - \mathbf{J}_z\mathbf{y} - \mathrm{i}\mathbf{x} \\ 0 &= \mathbf{J}_z\mathbf{x} - \mathbf{J}_x\mathbf{z} - \mathrm{i}\mathbf{y} \\ 0 &= \mathbf{J}_x\mathbf{x} + \mathbf{J}_y\mathbf{y} + \mathbf{J}_z\mathbf{z},\end{aligned} \tag{G.5}$$

where $\mathbf{J}_x$ etc are the $(2j+1)\times(2j+1)$ spin-j matrices introduced in §7.4.4 and $\mathbf{x}$ etc are the $(2j+1)\times(2j'+1)$ arrays of matrix elements that we seek to constrain. These are three linear homogeneous simultaneous equations for the three unknowns $\mathbf{x}$, etc. Unless the 3×3 matrix that has the $\mathbf{J}$ matrices for its elements is singular, the equations will only have the trivial solution $\mathbf{x} = \mathbf{y} = \mathbf{z} = 0$. One can demonstrate that the matrix *is* non-singular by eliminating first $\mathbf{x}$ and then $\mathbf{y}$. Multiplying the first equation by $\mathrm{i}\mathbf{J}_x$ and then subtracting the third, and taking the second from $\mathrm{i}\mathbf{J}_z$ times the first, we obtain

$$\begin{aligned}0 &= (\mathrm{i}\mathbf{J}_x\mathbf{J}_y - \mathbf{J}_z)\mathbf{z} - (\mathrm{i}\mathbf{J}_x\mathbf{J}_z + \mathbf{J}_y)\mathbf{y} \\ 0 &= (\mathrm{i}\mathbf{J}_z\mathbf{J}_y + \mathbf{J}_x)\mathbf{z} - (\mathrm{i}\mathbf{J}_z^2 - \mathrm{i})\mathbf{y}.\end{aligned} \tag{G.6}$$

Eliminating $\mathbf{y}$ yields

$$\begin{aligned}0 &= \{\mathrm{i}(\mathbf{J}_z^2-1)(\mathrm{i}\mathbf{J}_x\mathbf{J}_y-\mathbf{J}_z)-(\mathrm{i}\mathbf{J}_x\mathbf{J}_z+\mathbf{J}_y)(\mathrm{i}\mathbf{J}_z\mathbf{J}_y+\mathbf{J}_x)\}\mathbf{z}\\ &= \{-\mathbf{J}_z^2\mathbf{J}_x\mathbf{J}_y-\mathrm{i}\mathbf{J}_z^3+\mathbf{J}_x\mathbf{J}_y+\mathrm{i}\mathbf{J}_z \\ &\qquad +\mathbf{J}_x\mathbf{J}_z^2\mathbf{J}_y-\mathrm{i}(\mathbf{J}_x\mathbf{J}_z\mathbf{J}_x+\mathbf{J}_y\mathbf{J}_z\mathbf{J}_y)-\mathbf{J}_y\mathbf{J}_x\}\mathbf{z}.\end{aligned} \tag{G.7}$$

We can simplify the matrix that multiplies $\mathbf{z}$ by working $\mathbf{J}_z$ to the front. In fact, using

$$\begin{aligned}\mathbf{J}_x\mathbf{J}_z^2\mathbf{J}_y &= (\mathbf{J}_z\mathbf{J}_x-\mathrm{i}\mathbf{J}_y)\mathbf{J}_z\mathbf{J}_y = \mathbf{J}_z(\mathbf{J}_z\mathbf{J}_x-\mathrm{i}\mathbf{J}_y)\mathbf{J}_y-\mathrm{i}(\mathbf{J}_z\mathbf{J}_y+\mathrm{i}\mathbf{J}_x)\mathbf{J}_y\\ &= \mathbf{J}_z^2\mathbf{J}_x\mathbf{J}_y-2\mathrm{i}\mathbf{J}_z\mathbf{J}_y^2+\mathbf{J}_x\mathbf{J}_y\end{aligned} \tag{G.8}$$

and

$$\left.\begin{aligned}\mathbf{J}_x\mathbf{J}_z\mathbf{J}_x &= \mathbf{J}_z\mathbf{J}_x^2-\mathrm{i}\mathbf{J}_y\mathbf{J}_x\\ \mathbf{J}_y\mathbf{J}_z\mathbf{J}_y &= \mathbf{J}_z\mathbf{J}_y^2+\mathrm{i}\mathbf{J}_x\mathbf{J}_y\end{aligned}\right\} \Rightarrow \mathbf{J}_x\mathbf{J}_z\mathbf{J}_x+\mathbf{J}_y\mathbf{J}_z\mathbf{J}_y = \mathbf{J}_z(\mathbf{J}_x^2+\mathbf{J}_y^2)-\mathbf{J}_z \tag{G.9}$$

equation (G.7) simplifies to

$$\{\mathrm{i}\mathbf{J}_z(3-\mathbf{J}^2-2\mathbf{J}_y^2)+\mathbf{J}_x\mathbf{J}_y\}\mathbf{z}=0. \tag{G.10}$$

The matrix multiplying $\mathbf{z}$ is not singular, so $\mathbf{z}=0$. Given this result, the second of equations (G.9) clearly implies that $\mathbf{y}=0$, which in turn implies that $\mathbf{x}=0$. This completes the demonstration that the matrix elements of the dipole operator between states of well defined angular momentum vanish unless $|j-j'|\le 1$.

Appendix H: Restrictions on scattering potentials

The $\Omega_\pm$ operators of equation (12.3) require us to evaluate $\mathrm{e}^{\mathrm{i}Ht/\hbar}\mathrm{e}^{-\mathrm{i}H_0t/\hbar}$ as $t\to\mp\infty$. Since $\mathrm{e}^{\pm\mathrm{i}\infty}$ is not mathematically well defined, we must check we really know what $\Omega_\pm$ actually means.

We can determine $\Omega_\pm$ from equation (12.13) if it is possible to take the limit $t\to\mp\infty$ in the upper limit of integration. Hence the $\Omega_\pm$ operators will make sense so long as this integral converges when acting on free states.

Let's concentrate for a while on Ω_-, with $|\psi;0\rangle=\Omega_-|\phi;0\rangle$ telling us that $|\psi\rangle$ and $|\phi\rangle$ behave the same way in the distant future. Using equation (12.13), we have

$$|\psi;0\rangle=\Omega(t')|\phi;0\rangle+\frac{\mathrm{i}}{\hbar}\int_{t'}^{\infty}\mathrm{d}\tau\,U^\dagger(\tau)VU_0(\tau)|\phi;0\rangle. \tag{H.1}$$

To decide if the integral converges, we ask whether its modulus is finite (as it must be if $|\psi\rangle$ can be normalized). The triangle inequality $|\mathbf{v}_1+\mathbf{v}_2|\le|\mathbf{v}_1|+|\mathbf{v}_2|$ tells us that the modulus of an integral is no greater than the integral of the modulus of its integrand, so

$$\left|\int_{t'}^{\infty}\mathrm{d}\tau\,U^\dagger(\tau)VU_0(\tau)|\phi;0\rangle\right|\le\int_{t'}^{\infty}\mathrm{d}\tau\,\left|U^\dagger(\tau)VU_0(\tau)|\phi;0\rangle\right|. \tag{H.2}$$

Since $U(\tau)$ is unitary, the integrand simplifies to $|VU_0(\tau)|\phi;0\rangle|=|V|\phi;\tau\rangle|$ where $|\phi;\tau\rangle$ is the state of the free particle at time τ. If the potential depends only on position, it can be written $V=\int\mathrm{d}^3\mathbf{x}\,V(\mathbf{x})|\mathbf{x}\rangle\langle\mathbf{x}|$, and the integrand on the right of equation (H.2) becomes

$$\big|V|\phi;\tau\rangle\big|=\langle\phi;\tau|V^2|\phi;\tau\rangle^{1/2}=\left[\int\mathrm{d}^3\mathbf{x}\,V^2(\mathbf{x})|\langle\mathbf{x}|\phi;\tau\rangle|^2\right]^{1/2}. \tag{H.3}$$

What does this expression mean? At any fixed time, $|\langle\mathbf{x}|\phi;\tau\rangle|^2\,\mathrm{d}^3\mathbf{x}$ is the probability that we find our particle in a small volume $\mathrm{d}^3\mathbf{x}$. Equation (H.3) instructs us to add up these probabilities over all space, weighted by the square of the potential – in other words (with the square root) we calculate the rms $V(\mathbf{x})$ felt by the particle at time τ. As time progresses, the particle moves and we repeat the process, adding up the rms potential all along the line of flight.

Now $1=\langle\phi;\tau|\phi;\tau\rangle=\int\mathrm{d}^3\mathbf{x}\,|\langle\mathbf{x}|\phi;\tau\rangle|^2$, we can be confident that for any given value of τ the integral over $\mathbf{x}$ on the right of (H.3) will be finite. Since the integrand of the integral over τ is manifestly positive, convergence of the integral over τ requires that

$$\lim_{\tau\to\infty}\left[\int\mathrm{d}^3\mathbf{x}\,V^2(\mathbf{x})\,|\langle\mathbf{x}|\phi;\tau\rangle|^2\right]^{1/2} < \mathrm{O}(\tau^{-1}). \tag{H.3b}$$

We began our discussion of scattering processes by claiming that the real particle should be free when far from the target, so it's not surprising that we now find a condition which requires that the particle feels no potential at late times.

If we neglect dispersion, $|\langle \mathbf{x}|\phi;\tau\rangle|^2$ is just a function of the ratio $\boldsymbol{\xi} \equiv \mathbf{x}/\tau$ as the particle's wavepacket moves around. Assuming that the potential varies as some power r^{-n} at large radii, we have for large τ

$$\int \mathrm{d}^3\mathbf{x}\, V^2(\mathbf{x})\, |\langle \mathbf{x}|\phi;\tau\rangle|^2 \simeq \tau^{3-2n} \int \mathrm{d}^3\boldsymbol{\xi}\, V^2(\boldsymbol{\xi}) f(\boldsymbol{\xi}). \tag{H.4}$$

Hence, at late times the rms potential varies as $\sim \tau^{-n+3/2}$ and $\Omega\pm$ is certainly well defined for potentials that drop faster than $r^{-5/2}$. When dispersion is taken into account, we can sometimes strengthen this result to include potentials that drop more slowly – see Problem 12.4.

Unfortunately, the Coulomb potential does not satisfy our condition. We will not let this bother us too greatly because pure Coulomb potentials never arise – if we move far enough away, they are always shielded by other charges.