Atomic Physics

High-precision quantum systems and the interaction of light and matter

Dr Andrew Steane

April 10, 2002

Contents

Textbooks

I recommend:

Leighton, "Principles of Modern Physics," chapters 5,6,7,8. Still a good clear introduction, though is slightly muddled when dealing with exchange symmetry, so not recommended for that.

Woodgate, "Elementary Atomic Structure," excellent and thorough, though aimed at a level which may test the average student. Does not include X rays except very briefly.

Eisberg and Resnick, good summary, some very nice pictures.

Kuhn, "Atomic Spectra" is a good source of information on spectra, term diagrams, and is very good for X rays.

Feynman Lectures on Physics volume III chapter 19 is good for periodic table.

Haken and Wolf is accessible and fills in experimental details which are missing elsewhere.

Thorne, "Spectrophysics" goes beyond 2nd year level, has good experimental discussion.

Condon and Shortley, another standard reference.

1 Themes in Atomic Physics

If you take an everyday object such as a brick and crumble it up into little pieces, then the smallest piece which you could say is somewhat like a very small piece of brick is an atom. Once you try to go to something smaller, by looking inside the atom, suddenly what you find is nothing at all like a brick or anything from everyday life. It is a springy cloud of electrons surrounding an incredibly dense nucleus; it is an electric system with more resonances than a piano, some of which are the most precisely defined quantities in all of physics, and the classical physics of Newton's equations completely fails to give any sort of correct description of what is going on.

Atomic physics is the study of those properties of an atom which are associated primarily with the system of electrons, and with how they interact with light. It is of immense practical importance, from allowing us to probe the outer reaches of space to designing street lamps. It also plays a significant role in physics education by introducing many of the significant concepts of quantum mechanics in a precise way. These can then be taken up and applied in other areas such as particle physics and condensed matter physics. These concepts include not just quantisation of energy, but also more subtle ideas such as exchange symmetry and many-particle wavefunctions.

Before we try to understand the quantum mechanics of a solid material, it is useful first to understand a single atom. Interestingly, it is also useful to understand an atom first before we try to understand a proton, because the quantum theory of the electron in hydrogen is a lot easier than the quantum theory of the quarks in a proton. In fact, the hydrogen atom is the most precisely understood system in all of physics. For example, the difference in binding energies of the two most strongly bound states of the electron-proton system has been measured[1] to a fractional precision 1.8×10^{-14}. The Schrödinger equation is not precise to this degree because it does not account correctly for special relativity—yet even when we replace the Schrödinger equation by the relativistically correct Dirac equation (outside the syllabus) the theory is still not precise enough. There is a part per million contribution which arises, roughly speaking, from the nature of the vacuum, whose theoretical treatment is provided by the full relativistic quantum theory of the electromagnetic interactions, called Quantum Electrodynamics (QED). With QED, however, we do at last find agreement between experiment and theory to fantastic precision (up to order α^8, see later for the meaning of this). This agreement is one of the triumphs of modern physics—it allows us to be confident that our understanding of the natural world is built on accurate foundations. If the human race knows anything at all, it knows what happens when an electron orbits a proton.

Atomic physics allows us to explore our fundamental understanding of Nature in several other ways. Continuing the theme of precision, we can for example test for a variation in time of fundamental 'constants' such as the charge on the electron; such variation is certainly small since if a variation by one part in 10^{14} had taken place during the year 2000 then the atomic clocks in Paris would have noticed because comparisons of frequencies of different atoms are sufficiently precise[2]! The mass of the Z boson, normally associated with vast particle accelerator experiments, was measured in the basement of the Clarendon laboratory using two optical tables, associated lasers, polarizers, and a tube of bismuth gas in 1996 to a precision[3] of 5%, by detecting a transition in bismuth atoms which violates the principle that left-handed and right-handed fundamental processes must occur equally often. Such 'parity violating' processes are only caused by the so-called *weak* interaction which is mediated by the Z particle; this is the link between the observed optical effect (a polarization rotation with a special frequency-dependence) and the underlying fundamental physics.

---

[1] Niering *et. al*, Phys. Rev. Lett. **84**, 5496 (2000), also Holzworth *et. al* in Atomic Physics 17: Proceedings of the XVII International Conference on Atomic Physics (ICAP 2000) (American Institute of Physics, New York 2001). The result: $f_{1S--2S} = 2466\ 061\ 413\ 187\ 103\ \pm 46$ Hz.

[2] C. Salomon *at al.* in "Cold Atom Clocks," Atomic Physics 17: Proceedings of the XVII International Conference on Atomic Physics (ICAP 2000), pp 23-40 (American Institute of Physics, New York 2001).

[3] The experiment was precise to 2%; the theoretical extraction of the Z mass relied on highly complex atomic structure calculations which were precise to 5%.

Another direction is in the control of atomic motion using laser light. The methods of laser cooling which won the Nobel prize for Cohen-Tannoudji, Chu and Phillips in 1997 (see http://almaz.com/nobel/physics/physics.html) allow micro-Kelvin temperatures to be routinely achieved for clouds of atoms in laser traps, and developments built on this led in 1995 to the observation of Bose-Einstein condensation, a remarkable collective quantum phenomenon, in the groups of Cornell and Weiman and of Ketterle and collaborators (Science vol 269, p. 198 (1995); Physical Review Letters vol. 75, p.3969 (1995)), earning for them the Nobel prize in 2001. This led in 1997 to the observation of laser-like beams of matter, the so-called *atom laser*. This produces a wavelike beam of many sodium atoms (for example), where the de Broglie waves agree in phase and therefore have comparable brightness and frequency-precision to laser light, only they are waves of matter not light. This is wave-particle duality with a vengeance!

Laser manipulation of trapped atoms has also enabled the first steps in the field of *quantum information processing* or quantum computing to be taken. Here it is the precision of atomic physics methods, combined with the richness of atomic behaviour (compared with elementary particles) which allows the subtle and fragile techniques required for quantum computing to be achieved. What exactly is quantum computing? It is essentially a qualitatively new form of information processing which is based on the manipulation of quantum *entanglement*, the most subtle feature of quantum mechanics ... which unfortunately these lectures will barely touch on.

The interaction of light and matter is central to atomic physics because the primary way in which light is produced or absorbed in the natural world is by interactions betweens photons and electrons in atoms or molecules. This is in contrast to radio waves, for example, which interact with electrons moving in larger objects such as a metal wire. Almost all the light we see streaming to and fro in the world, bringing to us a sense of the leaves on the trees and the waves on the sea shore and the stars around us, was emitted by atoms or molecules in the leaf or the water or the star, rather than by something smaller or larger. In this sense, we may truly say that if you want to see what atoms look like, *look around you.*

Properties of atoms are basic tools of several other branches of physics—for example the speed of stars is deduced from the Doppler shift of the light they emit, and the reason it is possible to know the light wavelength emitted by the star is because we can recognise the signatures of the atoms in the outer layer of the star, which only emit at a set of characteristic wavelengths. The set is recognisable, but is found uniformly shifted in frequency. Observation of many other properties of stars also relies on arguments from atomic physics, for example we can deduce magnetic fields from the Zeeman effect. Experiments involving plasmas, such as fusion power research, similarly require detailed knowledge of the properties of free atoms and ions. Finally, to return to the humble street light; this is based on some simple properties of sodium atoms. Why do you think our streets are lit using an electrical discharge in a gas of sodium? ... familiarity with atomic physics and thermal radiation should enable you to figure this out.

The character of this area of physics is summed up for me by the adjective "exquisite". This suggests both the exquisite precision of the experiments which test our fundamental theories, the properties and significance of laser light, and also the exquisite subtlety of such experiments as those related to quantum entanglement and quantum computing.

1.1 Some mathematical notations

Throughout the lectures I will use the standard notation

$$\langle\psi|\,\hat{A}\,|\phi\rangle \equiv \int_{\text{all space}} \psi^*\hat{A}\phi\,dxdydz \tag{1}$$

where $\hat{A}$ is an operator and $\psi(x, y, z)$ and $\phi(x, y, z)$ are wavefunctions. The quantity $\langle \psi | \hat{A} | \phi \rangle$ is referred to as "the matrix element for operator $\hat{A}$ between states ψ and ϕ". It is also sometimes called a 'bracket'. The important thing about the notation $\langle \psi | \hat{A} | \phi \rangle$ is that we can use it equally for more general cases, such as when ψ and ϕ are wavefunctions for a system of more than one particle, in which case they are functions of several sets of coordinates, and the integral is over all space for all the sets of coordinates. The notation will also naturally include the case where the wavefunctions are functions of a degree of freedom other than position, such as spin for example. In the various cases, $\langle \psi | \hat{A} | \phi \rangle$ means whichever type of integral is appropriate to cover the complete range of the degrees of freedom (e.g. space, spin) which ψ and ϕ are functions of. Note that when $\psi = \phi$ we have $\langle \psi | \hat{A} | \psi \rangle$, i.e. the *expectation value* of operator $\hat{A}$ when the system is in the state ψ.

If there is any ambiguity about whether a quantity is a complex number or an operator, then I will put a hat on the operator, as $\hat{A}$ for example. A bold font operator such as $\mathbf{\hat{J}}$ is a *vector operator*, that is, an operator with three operator components $(\hat{J}_x, \hat{J}_y, \hat{J}_z)$.

Since we will be using the "matrix element" notation $\langle \psi | \hat{A} | \phi \rangle$ without always needing to write the integral explicitly, it won't always be necessary to use greek letters such as ψ and ϕ to label the wavefunctions. Typically, we are concerned with wavefunctions which are eigenstates of some particular observable such as energy or angular momentum. For example, suppose ψ_b is an eigenstate of operator $\hat{B}$:

$$\hat{B}\psi_b = b\psi_b \tag{2}$$

here b is the eigenvalue, and we have conveniently labelled ψ_b by its eigenvalue so as to keep track of which eigenstate it is. We might be interested in matrix elements such as

$$\langle \psi_b | \hat{A} | \psi_{b'} \rangle . \tag{3}$$

In such a case we will simplify the notation to

$$\langle \psi_b | \hat{A} | \psi_{b'} \rangle \equiv \langle b | \hat{A} | b' \rangle . \tag{4}$$

This keeps things uncluttered and makes sure we don't lose track of the prime on a subscript! Just to underline the point, I will state again what the notation means in terms of an integral:

$$\langle b | \hat{A} | b' \rangle = \int_{\text{all space}} \psi_b^* \hat{A} \psi_{b'} \, dxdydz. \tag{5}$$

SO, when the bracket symbol contains just eigenvalues inside the $\langle \cdots |$ and $| \cdots \rangle$ parts, we mean the integral is over the relevant eigenstates. I hope that is clear!

The following simple example is treated in the lecture: if ϕ_a is an eigenstate of $\hat{A}$ with eigenvalue a then it is simple to prove that

$$\langle a | \hat{A} | a' \rangle = \left\{ \begin{array}{ll} a & \text{if } a = a' \\ 0 & \text{otherwise} \end{array} \right\} \qquad \text{N.B. only for eigenstates of } \hat{A} \tag{6}$$

1.2 Atomic physics—some preliminaries

The atomic physicist counts "one, two, lots." That is to say, hydrogen and helium are both special cases, and the rest are understood in terms of concepts borrowed from them. Hydrogen is special because it can be solved exactly (ignoring quantum electrodynamics) and because the various relativistic effects are all of a similar order of magnitude, while in other atoms one of these tends to dominate (the spin-orbit interaction, more of this later). Helium is special because the

many-electron averaging which we use in atoms with many electrons is not so successful, and also various spin effects are all of the same order of magnitude (see chapter 11).

The following table gives the values of constants which will be much used and should be committed to memory to 3 significant figures. Their meaning will become clear as we go along.

Fine structure constant	α	$= \dfrac{e^2}{4\pi\epsilon_0 \hbar c}$	1/137.03599976
Rydberg energy	E_{R}	$= \frac{1}{2}m\alpha^2 c^2$	13.6056981 eV
Bohr radius	a_0	$= \hbar/m\alpha c$	0.529177249 Å
Bohr magneton	μ_B	$= e\hbar/2m$	$h \times 13.9962418$ MHz/mT
electron g factor	g	$= 2(1 + \frac{\alpha}{2\pi} - 0.328\frac{\alpha^2}{\pi^2} + \ldots)$	2×1.0011596

1.2.1 The role of classical and quantum mechanics

Quantum mechanics is the only way to understand atomic physics properly. However, because classical mechanics is so much more familiar, it a very useful tool for providing physical insight and estimates of various quantities such as energies, distances, etc. One should develop a feel for both.

1.3 Introducing the atom

The diagram in the lecture shows how to think about an atom. It is useful to pick hydrogen as a starting point, and then go on to a 'typical' atom having atomic number of order 10. Here are the order of magnitude sizes of some physical features which come up:

feature	hydrogen $(Z = 1)$	range for $Z = 1$ to 100	units
mean radius	0.79	0.5—2.7	Å
ionization potential	13.6	4—24	V
r.m.s. speed of an electron	0.01	0.001—1	c (speed of light)
electric field at mean radius	10	10—100	V/m
electric field at edge of nucleus	10^6	10^6—2×10^7	V/m
mean magnetic field experienced by outer electron	3	3—1000	T
magnetic field experienced by the nucleus	100	10—1000	T
electric dipole moment of the whole atom	0	0	Cm
magnetic dipole moment of the whole atom	10^{-23}	0—4×10^{-23}	J/T

The aim at this stage is to note the order of magnitude and the fact that for some properties a wide range is possible, while others do not vary much. In particular, notice that the electric and magnetic fields inside any atom are huge. The reasons for these values will be learned as we go on.

2 Hydrogen

It is common to begin the study of hydrogen using the Bohr theory. This is a theory which pre-dated a proper understanding of quantum mechanics, but which used some relevant concepts such as quantisation. While it has historical interest I feel it may confuse more than it illuminates at this stage, so I will ignore it altogether. What we need to do to understand hydrogen is write down and solve the Schrödinger equation for the system.

2.1 Schrödinger equation solution: Main features

Hydrogen consists of an electron orbiting a proton. By far the greatest contribution to the potential energy of the system is the simple Coulomb attraction. Therefore to an excellent approximation the Schrödinger equation for the system is

$$\frac{-\hbar^2}{2m}\nabla^2\psi_E - \frac{e^2}{4\pi\epsilon_0 r}\psi_E = E\psi_E, \tag{7}$$

where the first term $-\hbar^2\nabla^2/2m$ is the kinetic energy (equal to $\hat{p}^2/2m$ where $\hat{\mathbf{p}}$ is the momentum), and the second term is the potential energy which is negative because the particles attract one another.

In fact there are already a few subtleties which should not be ignored in writing equation (7). First, which particle are we talking about: the proton, the electron, or both? The answer is that we are considering the *relative coordinate* $\mathbf{r} = \mathbf{r_e} - \mathbf{r_p}$ and the net momentum in the centre of mass reference frame. Therefore the equation describes how the *relative* position and momentum of the particles behaves. Now, since the one-proton nucleus is about 1836 times heavier than the electron, it happens that the centre of mass is located to a good approximation at the nucleus, therefore we may regard the position vector $\mathbf{r}$ as essentially the position of the electron. However, it is not necessary to make that approximation. The relative coordinate and momentum behave *exactly* like those of a single particle in the given potential $V(r) = -e^2/4\pi\epsilon_0 r$, but having the *reduced mass*

$$m = \frac{m_e M}{m_e + M} \tag{8}$$

where m_e is the mass of the electron, and M is the mass of the nucleus (here, a proton). Note that m is always less than m_e, that is why it is called a reduced mass. It tends to m_e in the limit that M is large.

The reduced mass is an important consideration because we want to be precise, and indeed it becomes a major effect when we take an interest in the *difference* between between isotopes of the same element. Such atoms have the same number of electrons and protons, so they differ only in the number of neutrons in the nucleus. The difference in their energy levels is called an *isotope shift*. An example is given in problem set 4, and we will return to this in section 13.

The method of solution of equation (7) was discussed in the quantum mechanics lecture course. I will just give the highlights here.

First, since the potential is spherically symmetric (called a *central* potential) it is convenient to treat the problem in spherical polar coordinates. The Laplacian operator ∇^2, multiplied by $-\hbar^2/2m$ to express kinetic energy, is in spherical polar coordinates

$$\frac{-\hbar^2}{2m}\nabla^2 = \frac{-\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{\hat{\mathbf{L}}^2}{2mr^2} \tag{9}$$

where $\hat{\mathbf{L}}$ is the operator for angular momentum (defined by $\hat{\mathbf{L}} = \mathbf{r} \wedge \hat{\mathbf{p}}$). The first term is the contribution to kinetic energy coming from radial motion, i.e. motion in and out, while the second term is the kinetic energy of angular motion, i.e. rotation. (You might recognise the form $\hat{\mathbf{L}}^2/2mr^2$ as $\hat{\mathbf{L}}^2/2I$ where $I = mr^2$ is the moment of inertia.)

Multiplying the whole equation by r^2, and remembering that $\hat{\mathbf{L}}^2$ is a function of angular variables only, not radial ones, it is possible to see that the Schrödinger equation can be solved by the method of separation of variables. There exists a complete set of solutions having the form $\psi(r, \theta, \phi) = R(r)Y(\theta, \phi)$. The Y part you should recall from quantum mechanics lectures, it is an eigenstate of the operator $\hat{\mathbf{L}}^2$, with eigenvalue $l(l+1)\hbar^2$ where $l = 0, 1, 2, \cdots$ is a non-negative integer. It is also an eigenstate of the operator for the z-component of angular momentum $\hat{L}_z$, having eigenvalue $m\hbar$ where m is an integer in the range $-l \leq m \leq l$. Therefore we will label the Y by the subscripts l, m to remind us of this:

$$\hat{\mathbf{L}}^2 Y_{l,m}(\theta, \phi) = l(l+1)\hbar^2 \, Y_{l,m}(\theta, \phi) \tag{10}$$
$$\hat{L}_z Y_{l,m}(\theta, \phi) = m\hbar \, Y_{l,m}(\theta, \phi). \tag{11}$$

The radial part remains to be solved. It is a solution of the equation

$$\left(\frac{-\hbar^2}{2m} \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r) \right) R_E(r) = E R_E(r). \tag{12}$$

We have written $V(r)$ rather than the explicit form $-e^2/4\pi\epsilon_0 r$ as a reminder that the results so far are valid for any central potential. Now we use a standard trick, which is to write $R_E(r) = P_E(r)/r$. Note that $R_E(r)$ is the radial part of the wavefunction, while $P_E(r)$ is simply a related function which we have introduced for convenience[4]. Substituting it in, we find

$$\frac{-\hbar^2}{2m} \frac{d^2 P_E}{dr^2} + \left(V(r) + \frac{l(l+1)\hbar^2}{2mr^2} \right) P_E = E P_E. \tag{13}$$

Now let's guess the solutions. Since $V(r)$ is only linear in $1/r$, the so-called "centrifugal barrier" term $l(l+1)/r^2$ will dominate when $r \to 0$. Trying $P \sim r^k$ we obtain $k(k-1) = l(l+1)$ which has solutions $k = l+1$ or $(-l)$. Rejecting the latter as unphysical (tends to infinity at origin), we find

$$R_E(r) \propto r^l \ \text{ as } \ r \to 0. \tag{14}$$

This is already a useful observation: we see that the state of least angular momentum, $l = 0$, can be finite at the origin (it is proportional to $r^0 = 1$) while states of higher and higher angular momentum are flatter and flatter at the origin. This is important because *the origin is where the nucleus is*: we see that states of low angular momentum are more sensitive to the details of the nucleus, such as the fact that it is not really a point particle.

As $r \to \infty$ the potential and the centrifugal term fall off, so the kinetic energy dominates, and the solution is one which falls off exponentially as $\exp(-\kappa r)$ where $\kappa^2\hbar^2/2m = -E$ (this is a positive quantity since E is negative for bound states). The equation can be solved by a further change of function, followed by the method of series solution (see appendix). The energy eigenvalues are found to be

$$E = -\frac{me^4}{(4\pi\epsilon_0)^2 2\hbar^2} \frac{1}{(k+l+1)^2} \tag{15}$$

---

[4]Well, P doesn't have a direct physical interpretation, but $|P|^2$ does. It should be obvious to you, after a little thought, that $|P|^2 = r^2|R|^2$ is interesting because it is (when correctly normalised) the probability density for finding the electron at r, i.e. anywhere in the spherical shell of radius r, thickness dr.

where k is an integer in the range $k = 0, 1, 2, \ldots$, and the associated wavefunctions are also obtained. They should be looked up in the textbooks. The integer k is equal to the number of nodes (zero crossings) in the radial part of the wavefunction. The wavefunction of the ground state is worth remembering, it is:

$$\psi(r, \theta, \phi) = Ae^{-r/a_0} \tag{16}$$

where A is a normalisation constant which need not be memorized since it can be easily calculated. a_0 is the Bohr radius. Note that the left hand size of the expression emphasizes that the complete wavefunction over all space has been written down, in spherical polar coordinates. It so happens that this wavefunction does not have any θ and ϕ dependence: it is spherically symmetric, but other states are not necessarily spherically symmetric (those with $l > 0$ are not).

There are two important simplifications of equation (15). First we notice that the energies are all given by a constant divided by an integer squared. The integer $k + l + 1$ is called the *principal quantum number* and is indicated by the symbol n. We can now drop k since it is given by n and l, and we note that n and l have the ranges:

$$
\begin{aligned}
n &= 1, 2, 3, \ldots \infty \tag{17} \\
l &= 0, 1, 2 \ldots n - 1 \tag{18}
\end{aligned}
$$

Both n and l arise for any central potential, not just the $1/r$ potential, and $n - l - l$ is always the number of nodes in the radial wavefunction, but for potentials with some other dependence than $1/r$ the energy depends on both n and l. This case arises for atoms other than hydrogen.

The next simplification of equation (15) is to express the constant in a more physically meaningful way. The most useful insight is to consider the *fine structure constant* α defined by

$$\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} \tag{19}$$

where c is the speed of light. Consider the dimensions of this quantity: $e^2/4\pi\epsilon_0$ has the dimensions of an energy multiplied by distance, so $e^2/4\pi\epsilon_0 c$ has dimensions of energy multiplied by time (Joule-seconds in SI units), therefore α is dimensionless. The fine structure constant is absolutely central to the quantum physics of electromagnetic interactions: it is a combination of fundamental constants which compares electromagnetic quantities with Planck's constant. The value of α is 7.29735×10^{-3}, and it is easy to remember this as $\alpha \simeq 1/137$ which is accurate to one part in 4000.

You can show (see problems) and it is easy to remember that the mean speed (strictly, the r.m.s. speed) of the electron in the ground state of hydrogen is equal to αc. Furthermore, the potential energy is equal to minus twice the kinetic energy (this is called the virial theorem) and therefore the total energy equals minus the kinetic energy:

$$E = -\frac{1}{2}m\alpha^2 c^2 \tag{20}$$

for the ground state. You should confirm that this agrees with equation (15), that is, it is just a convenient way of expressing the answer obtained by solving the Schrödinger equation. We will see later (section 11) that the expression in terms of αc arises naturally out of a relativistically correct treatment (i.e. one valid for speeds of order c).

Finally, then, all the energy levels of hydrogen are given by

$$E_n = -\frac{1}{2}m\alpha^2 c^2 \frac{1}{n^2}. \tag{21}$$

The energy $\frac{1}{2}m_e\alpha^2 c^2$ is called the Rydberg energy, where m_e is the standard mass of a free electron ($\sim 9.10939 \times 10^{-31}$ kg). Note that by convention the Rydberg energy is defined in terms of the full electron mass, not the reduced mass, so the ionisation energy of the hydrogen ground state is not exactly 1 Rydberg but in fact slightly less than 1 Rydberg (even before relativistic corrections are taken into account).

The ground state wavefunction is found to be

$$\psi_1(r,\theta,\phi) = \frac{1}{(\pi a_0^3)^{1/2}} e^{-r/a_0} \tag{22}$$

where the constant $a_0 = 4\pi\epsilon_0\hbar^2/me^2$ is called the Bohr radius. The Bohr radius is important because it gives the size of the hydrogen atom—that is, we know the wavefunction does not have a sharp cut-off, but a_0 gives the distance scale over which the wavefunction is falling off as r increases. It is useful to remember a_0 in terms of α, which can be done by noting that an electron moving at speed αc at a distance of a_0 from some point has angular momentum $\hbar$ about that point, hence $a_0 m\alpha c = \hbar$, from which

$$a_0 = \frac{\hbar}{m\alpha c} \simeq 0.5 \text{ Å} \tag{23}$$

2.2 Comment

What is the meaning of n? Out of the smoothness of the Schrödinger equation an integer, a discrete thing, has emerged. Why? Because out of all the continuous infinity of candidate solutions to the Schrödinger equation, only a discrete set behave correctly as $r \to 0$ and $r \to \infty$. The wavefunction is like a standing wave which has to have a specific form at these two 'boundaries', and in between it can oscillate up and down an integer number of times. To be specific, it passes through zero $n - (l + 1)$ times. The reason we choose to call n the 'principal quantum number' rather than $n - (l + 1)$ is to draw attention to the fact that the energy eigenvalue depends primarily on n. Indeed, for the $1/r$ potential, by a mere chance (actually by a hidden symmetry) the energy depends only on n and not on l at all. However, n is primarily associated with the eigenfunctions, rather than the eigenvalues, and an equivalent quantum number will appear whenever we have a central potential, of whatever dependence on r. The eigenvalues will then still be related to n and l, but not by a simple formula.

2.3 Orbital angular momentum notation

For historical reasons, the orbital angular momentum quantum number l is associated with a letter, according to the following table:

l =	0	1	2	3	4	5	...
	s	p	d	f	g	h	...

(the letters arose from observations on spectra: 's' for sharp, 'p' for principal, 'd' for diffuse, 'f' for fundamental (for some reason!), then g,h,....)

Using this notation the radial part of the wavefunctions of hydrogen can be indicated by giving the values of n and l in the form

n	l	notation	radial wavefunction
1	0	1s	$(\pi a_0^3)^{-1/2} \exp(-r/a_0)$
2	0	2s	$(8\pi a_0^3)^{-1/2} (1 - r/2a_0) \exp(-r/2a_0)$
2	1	2p	$(8\pi a_0^3)^{-1/2} (r/2\sqrt{3}a_0) \exp(-r/2a_0)$
3	0	3s	etc.
3	1	3p	...look them up in Leighton/Woodgate/...!
3	2	3d	
etc.			

In this table, the radial part of the wavefunction has been normalised such that $\int_0^\infty R^* R \, 4\pi r^2 dr = 1$, but note that other ways of normalising are possible, depending on how the normalization of the full wavefunction $R(r)Y(\theta, \phi)$ is apportioned between the radial and angular parts—a matter of convention and convenience.

2.4   Some classical estimates

We have now developed a good understanding of the quantum theory of hydrogen. It is useful to add to this some simple calculations which are based on a rough-and-ready combination of classical and quantum physics. For example, consider a classical particle of charge $-e$ moving at speed v in a circular orbit of radius r about a fixed charge $+e$. The force is $e^2/4\pi\epsilon_0 r^2$ so Newton's law "$F = ma$" gives

$$mv^2 = e^2/4\pi\epsilon_0 r. \tag{24}$$

This shows that the potential energy is equal to minus twice the kinetic energy, and is also a relationship between v and r. Next (this is the quantum part) let's assume the orbital angular momentum mrv is equal to $\hbar$. Why should it equal $\hbar$? There is no reason, but if the system is in a low-lying quantum state this is liable to be the order of magnitude of the angular momentum. Putting $mrv = \hbar$ into (24) we find

$$v = e^2/4\pi\epsilon_0\hbar = \alpha c$$

and the total energy is $-(1/2)m\alpha^2 c^2$ which is exactly equal to the Rydberg, i.e. the energy of the ground state predicted by the full theory. That is nice, but don't forget the true quantum ground state doesn't have any orbital angular momentum, so the 'classical orbit' picture isn't completely successful. [Problem: use this method to examine the circular orbits of angular momentum $l\hbar$, and show that the quantum and classical results agree fully in the limit of large l and large n.]

You should be aware that the classical solution of two point-like particles attracting one another leads to elliptical orbits. Such orbits are called Bohr orbits, and in fact Bohr arrived at the correct energy levels by an argument like the one above, but made precise rather than 'rough and ready', in that his criteria were to make specific assumptions about quantisation of angular momentum, and to ensure that the limit of high n agreed with classical physics (the *correspondence principle*). In the limit of large n, certain superpositions of quantum energy eigenstates give probability densities which look like a lump of high probability (a wavepacket) going around the nucleus in an elliptical orbit, and thus the quantum and classical descriptions merge. There is one possible area for confusion, however: the classical orbit with the largest angular momentum for given energy is the circular one, while the quantum wavefunction with the *least* angular momentum for a given energy is the spherically symmetric one.

2.5 How to remember hydrogen

The ground state Bohr orbit has	$v/c =$	α
The ground state energy is	$E_1 =$	$-T_1$
	$=$	$-\frac{1}{2}m\alpha^2 c^2$
an electron orbiting at speed αc at the Bohr radius		
has one unit of angular momentum, so	$mva_0 =$	$\hbar$
from which	$a_0 =$	$\hbar/(m\alpha c)$
and more generally	$\langle r \rangle \propto$	n^2

2.6 Hydrogen-like systems

Our understanding of hydrogen can be applied almost unchanged to any system involving a pair of oppositely charged particles bound to each other. We simply adjust for the charge and mass of the particles. Any atom can be converted into a hydrogen-like ion by removing all but the last electron. Then if the atomic number is Z we have a nucleus of charge Z and the reduced mass of the electron is slightly changed since the nucleus is now heavier. We can immediately solve the problem without significant further work by looking carefully at the Schrödinger equation

$$\frac{-\hbar^2}{2m}\nabla^2\psi_E - \frac{Ze^2}{4\pi\epsilon_0 r}\psi_E = E\psi_E, \tag{25}$$

and reasoning as follows: the only change compared to hydrogen is the Z which multiplies $e^2/4\pi\epsilon_0$ in the equation (the fact that the reduced mass will not be the same as for hydrogen is already implicitly allowed for since the m in the equation is the reduced mass). When we solve for the behaviour, everywhere in the maths where previously $e^2/4\pi\epsilon_0$ occured, there will now instead be $Ze^2/4\pi\epsilon_0$. It follows that we can just write down the answer, using the rule that wherever you see e^2, now put Ze^2. When we write the solution in terms of the fine structure constant α for convenience, it means that wherever previously we had α we now have $Z\alpha$. Here is what we find:

The ground state Bohr orbit has	$v/c =$	$Z\alpha$
The ground state energy is	$E_1 =$	$-T_1$
	$=$	$-\frac{1}{2}mZ^2\alpha^2 c^2$
from the expression for a_0, we deduce	$\langle r \rangle \propto$	n^2/Zm

Notice that these results make physical sense: a highly charged nucleus pulls the electron in closer to it, and it can move at higher speeds while remaining bound.

The case of deuterium has $Z = 1$ but the mass of the nucleus is twice that of hydrogen [Problem: how many times precisely?: Look it up in a book of physical constants], therefore the reduced mass of the electron changes, and all the energy levels are slightly shifted—see problem set 4.

Two other interesting cases are positronium, and muonic atoms. Positronium is the system of an electron and a positron orbiting one another. The two particles will annihilate each other after about a microsecond (or 10 nanoseconds, see Leighton p. 624) but up until that time they form a hydrogen-like system with the important property that both particles are truly point-like: there is no internal structure, in contrast to the quarks inside the proton in hydrogen. This means the properties can be calculated even more precisely than for hydrogen, and one can test in particular the theory of the Coulomb interaction down to very small distances, small compared to the size of a proton, by very precise measurements of the spectrum.

If atoms are placed in a beam of muons from a particle accelerator, occasionally an electron in the atom will be knocked out and a muon captured. The muon is 206.8 times heavier than an electron,

so its wavefunctions have a much smaller Bohr radius. It soon decays to the ground state and then owing to the small Bohr radius it sits inside the majority of the charge distribution provided by the electrons. Therefore it experiences almost the whole nuclear charge and its energies are approximated by the hydrogen formula with $Z\alpha$ where Z is almost the full charge of the nucleus, and m is now the reduced muon mass. The only thing is, the muon orbit can be so small that it spends most of its time right inside the nucleus! Therefore the whole calculation needs to be reworked, (we can no-longer assume a point-like nucleus) and measurements on muonic atoms yield much information on nuclear properties.

Finally, another important hydrogen-like system is antihydrogen (!): a positron bound to an anti-proton. Accurate comparisons between the spectra of hydrogen and antihydrogen are just becoming feasible in laboratory experiments, and needless to say they give direct evidence of the nature of the fundamental symmetries of particles and their anti-particles. Antihydrogen, being the anti-atom which is easiest to fabricate, will also provide the first experimental test to ensure that gravity acts the same on particles and anti-particles. [Problem: why can't we test that using anti-protons or positrons which are more readily available from particle accelerators?]

2.7 Main points

The main points of this section have been:
Schrödinger equation for a central potential; $RY_{l,m}$; n, l quantum numbers and their meaning; $E_{\mathrm{R}} = -\frac{1}{2}m\alpha^2 c^2$; reduced mass (e.g. positronium); Bohr radius; the form of the wavefunctions; the emission spectrum; scaling with Z; the semi-classical picture.

2.8 Appendix on series solution of hydrogen equation, off syllabus

This calculation is "off syllabus" in the sense that it would not normally be set as an atomic physics question in finals, but as a part of mathematical methods in physics it is a standard example of the series solution of a differential equation, which you should know how to handle.

The radial Schrödinger equation for a $1/r$ potential will be solved, using previous knowledge of the angular part of the solution. This is a long calculation, and the detailed working between steps will involve further algebra which is not set down here. In other words, there is no *quick* derivation of the familiar $E \sim 1/n^2$ answer. The derivation will nevertheless illuminate some important features of the solution wavefunctions, and the meaning of the quantum number n.

With the clue that the solution falls off exponentially at large r, adopt the dimensionless variable

$$\rho = \sqrt{-2mE}\, r/\hbar \quad [\text{ NB } E < 0 \text{ so the sqrt is real}$$
$$\text{and let} \quad P_E = e^{-\rho}u \tag{26}$$

$$\implies 0 = \frac{d^2u}{d\rho^2} - 2\frac{du}{d\rho} + \left[\left(\frac{2m\alpha^2c^2}{-E} \right)^{1/2} \frac{1}{\rho} - \frac{l(l+1)}{\rho^2} \right] u \quad [\text{ using } V(r) = e^2/4\pi\epsilon_0 r$$

$$\text{Try} \quad u = \rho^{l+1} \sum_{k=0}^{\infty} C_k \rho^k \quad [\text{ remember } U \sim r^{l+1} \text{ at } r \to 0$$

$$\implies \frac{C_{k+1}}{C_k} = \frac{-(-2m\alpha^2c^2/E)^{1/2} + 2(k+l+1)}{(k+l+2)(k+l+1) - l(l+1)} \quad [\text{ after some work!}$$

$$\implies E = -\frac{\frac{1}{2}m\alpha^2c^2}{(k+l+1)^2} \quad [\text{ for the series to terminate}$$

$$= -\frac{E_1}{n^2} \quad [\textbf{ Hydrogen energy levels}$$

$$\text{where} \quad n = k+l+1, \quad k = 0,1,2,\ldots; \quad l = 0,1,2,\ldots$$

$$(27)$$

3 Grating spectroscopy and the emission and absorption spectrum of hydrogen

The main experimental techniques to investigate the structure of free atoms are emission and absorption spectroscopy. In absorption spectroscopy, light from a source with a continuous spectrum over a wide enough range is directed through a vapour cell containing a vapour of the atoms to be investigated. The spectrum of the transmitted light is measured, and found to contain a continuous background interrupted by narrow absorption lines at those wavelengths which the atoms can absorb. In emission spectroscopy atoms in a vapour are excited, typically by means of electron bombardment in a discharge, and the spectrum of the emitted light is measured. It is found to consist of a set of narrow lines.

The workhorse instrument for examining the gross structure of the spectrum is the diffraction grating. This is especially so for emission spectroscopy, since many of the wavelengths for neutral atoms are in the range near-ultra-violet to near-infra-red, which can be investigated using gratings. In absorption spectroscopy, wavelengths are generally shorter (for a simple reason we will examine in a moment) and diffraction gratings with sufficient reflectivity may not be available. A quartz prism spectrometer can be used in the ultra-violet region.

The details of emission and absorption spectroscopy are well described in two experiments on the practical course: I recommend obtaining the manuscripts and doing the experiments. The main points to consider in any form of spectroscopy are resolution, i.e. the smallest separation between spectral features which can be resolved by the instrument, and signal to noise ratio, in this case determined by the overall transmission from source to detector.

3.1 Main points for use of grating spectrograph

A diagram was given in the lecture. The main considerations for correct use of the instrument are

- Is full use made of the light available from the source? A condenser lens is used to focus the light onto the entrance slit. Note, however, that too narrow a focus will not help since the beam after passing through the slit expands again, and much of it will miss the collimation lens altogether if the focus was too tight.

- Is full use of the width of the grating made? The whole width of the grating must be illuminated to obtain the best resolution. The condenser lens should be chosen appropriately to bring this about.

- Is the light properly collimated at the grating? The collimation lens must be carefully adjusted to one focal length from the slit; this can be done by observing the transmitted light through a telescope which is known to be accurately focussed at infinity.

- Is a sharp image of the slit formed in the detection plane? The imaging lens is adjusted while the slit image is observed.

- Is the slit as narrow as it usefully can be? The minimum useful slit size is determined by the imperfect sharpness of the image owing to diffraction effects associated with the finite size of the diffraction grating and collimation/imaging optics. Below this limit, further reduction of slit size merely causes the image to be broader and less bright.

- Is stray light reaching the detector reduced to a minimum?

3.2 Resolution

The resolution of all optical spectroscopy instruments can be understood by one simple result:

The minimum resolvable wavenumber difference is equal to the inverse of the longest path difference over which interference occurs in the instrument.

(Wavenumber is one over wavelength, $\tilde{\nu} = 1/\lambda$).

For example, for a diffraction grating of width w used at normal incidence, if one of the diffracted orders for some wavelength λ is found to be at angle θ, then the longest path difference for interference is $w \sin \theta$, therefore the minimum resolvable wavenumber difference is

$$\delta\tilde{\nu} = \frac{1}{w \sin \theta}$$

Combining this with the fact that the n'th order diffraction occurs at angle θ given by $d \sin \theta = n\lambda$ for a grating of spacing d we have

$$\delta\tilde{\nu} = \frac{d\lambda}{wn} \tag{28}$$

$$\Rightarrow \quad \frac{\delta\tilde{\nu}}{\tilde{\nu}} = \frac{1}{nN} \tag{29}$$

where $N = w/d$ is the number of lines in the grating.

Proof of the above. As long as the general principle quoted above is properly understood, it is perfectly correct to quote it and therefore there is no strict need for further comments on the resolution of a diffraction grating. However, since the general principle may be unfamiliar, I will derive the diffraction grating resolution another way to confirm the answer. This will also add further physical insight, which is always valuable.

The resolution of the diffraction grating is limited by diffraction effects which arise since the grating has finite width. Treat the whole grating as a single (large) 'aperture', illuminated by plane waves at normal incidence. The light transmitted into the n'th order peak for wavelength λ emerges at angle θ to the normal. At angle $\theta + \Delta\theta$ the transmitted intensity falls to zero, where $\Delta\theta$ is found from the condition that for this angle the wavelets from the centre of the grating are $(p + 1/2)\lambda$ out of step with those from the edge, where p is an integer. (Then all the wavelets from across the grating sum in pairs to zero). This condition is:

$$\frac{w}{2} \sin(\theta + \Delta\theta) = \left(p + \frac{1}{2}\right)\lambda$$

where the (large) integer p is given by

$$\frac{w}{2} \sin \theta = p\lambda.$$

Expanding $\sin(\theta + \Delta\theta) \simeq \sin(\theta) + \cos(\theta)\Delta\theta$ for $\Delta\theta \ll 1$ and using (3.2) we obtain

$$\Delta\theta = \frac{\lambda}{w \cos \theta}. \tag{30}$$

A suitable criterion for a nearby wavelength $\lambda + \Delta\lambda$ to be resolved is that the maximum of its diffraction pattern falls at the first minimum of the diffraction pattern for λ, i.e. $\Delta\theta$ away from θ.

Therefore the minimum resolvable wavelength difference is given by

$$\Delta\lambda \quad = \quad \frac{d\lambda}{d\theta}\Delta\theta \tag{31}$$

$$= \quad \lambda\frac{d}{nw} \tag{32}$$

$$\text{using} \quad d\sin\theta \quad = \quad n\lambda \quad \Rightarrow \quad \frac{d\lambda}{d\theta} = \frac{d\cos\theta}{n} \tag{33}$$

$$\Rightarrow \quad \frac{\Delta\lambda}{\lambda} = \frac{\delta\tilde{\nu}}{\tilde{\nu}} \quad = \quad \frac{1}{nN} \tag{34}$$

as before.

3.3   Usefulness of both emission and absorption methods

Emission spectroscopy can be used to examine all the transitions of any atom (within technical limits), so it might be asked whether absorption spectroscopy adds anything. Both are useful because the spectra of atoms (and of anything else more complicated than an atom) are extremely complex and hard to interpret. One is faced with a plethora of lines and a lot of ingenuity is required to make sense of the data. The great boon of the absorption method is that a simpler spectrum is obtained: if the atoms are at typical temperatures (below a few hundred degrees C) then they will almost all be in the ground state, therefore only transitions from the ground state are observed. This enables the ground state transitions to be identified in the emission spectrum, which is the first step towards interpreting it.

3.4   The spectrum for hydrogen

For hydrogen the energy level formula $E_n = -E_R/n^2$ enables us to calculate the transition frequencies: for a transition from $n = j$ to $n = k$ the frequency is

$$\nu = \frac{|E_k - E_j|}{h} = \frac{E_R}{h}\left|\frac{1}{k^2} - \frac{1}{j^2}\right|.$$

The absorption spectrum was shown in the lecture: it consists of a sequence of lines of increasing frequency (decreasing wavelength), the longest wavelength is 122 nm (make sure you confirm this by calculation, to be sure you know where it comes from), the lines get closer and closer at shorter wavelengths, up to a 'series limit' at 91 nm (confirm this yourself too).

The emission spectrum has lines at the same wavelengths as the absorption spectrum, and also further lines at the frequencies of all the other transitions. These occur as a set of identifiable groups or series, each being the set of transitions to a given lower level. The series limit of each group enables the energy of this lower level to be deduced immediately. Similar series occur in more complicated atoms, where this observation is a useful pointer to aid the interpretation process.

The transition $n = 3$ to $n = 2$ is the beautiful red "Balmer-alpha" line of wavelength 656 nm.

3.5   What is going on when atoms emit and absorb light

We will discuss the coupling between atoms and light more thoroughly towards the end of the course. For now, I would simply like to give some guides to your physical intuition.

When an atom (or any other physical system) is in an energy eigenstate, it is in a certain important sense a *static* system. It is correct to picture the wavefunction ψ as oscillating in time, and the electron de Broglie waves as moving from place to place, but in an energy eigenstate all this wave motion always has the overall result that the probability distribution $|\psi|^2$ is completely static. Therefore when you think about an atom in one of its energy eigenstates, if you want to contemplate the net charge distribution within the atom, you should imagine something completely fixed, as unmoving as if it were set in concrete.

When an atom emits or absorbs light, on the other hand, it is temporarily in a superposition of energy eigenstates and the charge distribution oscillates. The oscillation continues for as long as it takes the atom to complete the absorption or emission process. The oscillation frequency is the frequency of the emitted or absorbed light.

Another interesting point is that when it comes to absorption and emission, a given atom is a completely different 'beast', depending on what state it is in. For example, hydrogen in the ground state has resonant frequencies at 2466 THz, 2923 THz, 3083 THz, etc., and can only absorb, while hydrogen in the 4s state has resonant frequencies at 3083 THz, 617 THz, 160 THz at which it can emit, and 74 THz, 114 THz, etc. at which it can absorb.

3.6 Main points

The main points of this chapter have been:
basic features of experimental method for emission and absorption spectroscopy; use of and resolution of the diffraction grating spectrometer; physical picture of absorption and emission process.

4 How quantum theory deals with further degrees of freedom

Before we can go on to Helium, and further features of hydrogen, it is necessary to review how quantum mechanics works when more than one particle is under consideration, and to introduce the concept of spin. This chapter addresses these concepts. The aim is to clarify some ideas which will have been met already in a course on quantum mechanics, but which typically need another look before the student feels comfortable with the material.

4.1 Multi-particle wavefunctions

Let's begin with known territory. We typically begin to learn quantum mechanics by treating a single particle in 1 dimension. We then have wavefunctions $\psi(x, t)$ which are functions of time and one other parameter: the position x of the particle along the x direction. In going to three dimensions, the change is to a function of time plus three other parameters: the 3 position coordinates, and when operators such as the Hamiltonian are written down, the dependences on all 3 coordinates and their associated momenta are included. Sometimes a 3D wavefunction is a product of three 1-dimensional ones, such as $\psi(x, y, z) = \sin(k_x x) \sin(k_y y) \sin(k_z z)$, but it does not have to be. For example a particle could be in a state such as

$$\psi(x, y, z) = (\sin(k_x x) \sin(k_y y) + \cos(k_x x) \cos(k_y y)) \sin(k_z z)$$

Here, there is no way to write the x and y dependence as a product of a function of x times a function of y.

In going from 1 to 3 dimensions, two new parameters of the particle were introduced (its y and z position). We took this into account simply by allowing the wavefunction to depend on these further parameters.

Next let us consider a pair of particles. The correct physical intuition is to see that one particle can do its thing without necessarily any relation to the other particle, or their motions can be correlated (as they typically will be if the particles interact). Both statements are of the same kind as those that can be said about the different coordinates of a single particle. This is the clue to how the maths works. The mathematical description simply treats the coordinates of the second particle as more parameters to be taken into account. We treat the two-particle system as a single physical system with 6 parameters. The wavefunctions ψ are functions of $x_1, y_1, z_1, x_2, y_2, z_2$ and they may or may not be in the form of a product of single-particle wavefunctions. The quantity $|\psi|^2 dx_1 dy_1 dz_1 dx_2 dy_2 dz_2$ is the probability to find the first electron between x_1, y_1, z_1 and $x_1 + dx_1, y_1 + dy_1, z_1 + dz_1$ *and* the second electron between x_2, y_2, z_2 and $x_2 + dx_2, y_2 + dy_2, z_2 + dz_2$.

Here as an example, it is a wavefunction of two particles in one of the energy eigenstates of a cubical box of side a:

$$\psi_{gf}(x_1, y_1, z_1, x_2, y_2, z_2) = \frac{8}{a^3} \sin(\pi x_1/a) \sin(\pi y_1/a) \sin(\pi z_1/a) \sin(2\pi x_2/a) \sin(\pi y_2/a) \sin(\pi z_2/a)$$

In this example the wavefunction is a product of single-particle wavefunctions, therefore it is legitimate to regard the particles as moving independently of each other: the first is in the ground state of the box, the second is in the first excited state of x motion. The total energy of the pair

is $E =???$. In order to bring out the product form, let us define

$$g(x, y, z) = \left(\frac{8}{a^3}\right)^{1/2} \sin(\pi x/a) \sin(\pi y/a) \sin(\pi z/a)$$

$$f(x, y, z) = \left(\frac{8}{a^3}\right)^{1/2} \sin(2\pi x/a) \sin(\pi y/a) \sin(\pi z/a)$$

then we have

$$\psi_{gf}(x_1, y_1, z_1, x_2, y_2, z_2) = g(x_1, y_1, z_1) f(x_2, y_2, z_2).$$

It will be useful later on to notice that if we switch around which particle is in which state, we get back a different state of the same energy:

$$\psi_{fg}(x_1, y_1, z_1, x_2, y_2, z_2) = f(x_1, y_1, z_1) g(x_2, y_2, z_2)$$
$$= \frac{8}{a^3} \sin(2\pi x_1/a) \sin(\pi y_1/a) \sin(\pi z_1/a) \sin(\pi x_2/a) \sin(\pi y_2/a) \sin(\pi z_2/a)$$

Two different states having the same energy are said to be *degenerate*. A centrally important property of such states is that any combination of two or more degenerate states will also be an energy eigenstate having the same energy as the degenerate states under consideration. (Proof: if $H\psi_{gf} = E\psi_{gf}$ and $H\psi_{fg} = E\psi_{fg}$ then $H(\alpha\psi_{gf} + \beta\psi_{fg}) = \alpha E\psi_{gf} + \beta E\psi_{fg} = E(\alpha\psi_{gf} + \beta\psi_{fg})$.) Therefore a further state which is also an eigenstate of energy of two particles in a box is

$$\frac{1}{\sqrt{2}}\left(\psi_{gf} + \psi_{fg}\right) = \frac{1}{\sqrt{2}}\left(g(x_1, y_1, z_1) f(x_2, y_2, z_2) + f(x_1, y_1, z_1) g(x_2, y_2, z_2)\right).$$

Note that for this case there is no meaningful way to assign a well-defined state to either particle on its own. Such states, which cannot be written as a single product of single-particle states, are said to be *entangled*.

4.2 Spin

Let us return again to the case of a single particle, but now introduce a new physical property, namely the intrinsic angular momentum of the particle, called its *spin*. We can regard spin either as a physical phenomenon which is met in the natural world when we do experiments, or as a property which arises naturally out of the relativistic (Lorentz covariant) formulation of quantum mechanics, discovered by Dirac. The main points to note here are that

1. Spin is a form of angular momentum.

2. Every fundamental particle has a spin, whose absolute magnitude (which might be zero) is an intrinsic property of the particle (in addition to its charge and mass etc.)

3. It is a new property in addition to, and quite unrelated to, the position of the particle.

The third point implies that the wavefunctions we have written down up till now, which are functions only of position, simply fail to account for spin altogether. They have to be extended.

The extension is similar to the ones we used above for going from 1 dimension to 3, or from 1 particle to 2, i.e. we just introduce a new parameter, but the difference is the parameter is not a position parameter. The new parameter lives not in position space but in 'spin space', an abstract mathematical space, and it has properties which reflect the properties of angular momentum which will be discussed in the next chapter.

The best way to get used to this is to consider the simplest possible case (other than spin zero), which is 'spin-half'. The spin state can here be modelled as a 2-component vector with complex coefficients. Thus there are two 'basis states' which can be written

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and a general spin state can be written as a linear combination of these. Including the position state as well, we can write some general state of a single particle in the form

$$\psi(\text{spatial and spin}) = af(x,y,z) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + bg(x,y,z) \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Note that the complete state does not have to be a product of spatial and spin parts, though it is possible to have states of that form.

Finally, another common notation for spin states is

$$|\uparrow\rangle \;\; = \;\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{35}$$

$$|\downarrow\rangle \;\; = \;\; \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{36}$$

In this notation we could write

$$\psi(\text{spatial and spin}) = af(x,y,z) |\uparrow\rangle + + bg(x,y,z) |\downarrow\rangle \,.$$

Expressions of this form are clear, and therefore can be used, but students who prefer to use Dirac notation throughout should adopt the form

$$|\text{spatial and spin state}\rangle = a\,|f\rangle \otimes |\uparrow\rangle + b\,|g\rangle \otimes |\downarrow\rangle$$

where $|f\rangle$ and $|g\rangle$ are understood to be spatial states, and I have included the tensor product symbol $\otimes$ for completeness, it may be left out in practice.

4.3   Main points

The main points of this chapter have been:
$\psi(x_1, y_1, z_1, x_2, y_2, z_2)$, its possible forms and its interpretation; spin as a new degree of freedom; spin-half state space spanned by $|\uparrow\rangle$, $|\downarrow\rangle$.

5 Angular momentum in quantum systems

Angular momentum features prominently in atomic physics. We have already seen that the eigenstates of energy for a particle in any central field are states of well-defined orbital angular momentum. It is correct to think about angular momentum by considering that an electron with non-zero orbital angular momentum is travelling around the nucleus. A classical picture of a circular or elliptical orbit gives some reasonable insight, and indeed if we examine the probability current density of the electron, it is found to be flowing around the nucleus. For zero orbital angular momentum this picture does not work any more. The electron is then not moving around the nucleus, but only in and out. The net flow in or out is zero, and this can be understood as a wave going in combined with a wave coming out; this explains how the electron can have kinetic energy without succeeding in going anywhere.

Electrons, in common with most fundamental particles, also have an intrinsic angular momentum, called 'spin', which never changes. In this respect it is an intrinsic property like the charge or mass of a particle. However, angular momentum is not a scalar but a vector quantity, and whereas the *size* of the spin does not change, its *direction* can. Therefore the spin as well as the orbital angular momentum must be treated by a careful quantum mechanical analysis.

The basic principles of how angular momentum behaves were introduced in the quantum mechanics lectures. Here we will just quote the answers as a simple set of rules. If you want to see the proofs, you should consult a quantum mechanics textbook, and I recommend *Principles of Quantum Mechanics* by R. Shankar (pub: Plenum Press).

Rules for angular momentum: (A) a single angular momentum

1. Angular momentum is represented by a vector operator, such as $\hat{\mathbf{J}}$. The size of the angular momentum is best understood by considering not the length of the angular momentum vector, but the length squared. This is because the operator $\hat{\mathbf{J}}^2 = \hat{\mathbf{J}} \cdot \hat{\mathbf{J}}$ is Hermitian and well-behaved, while any possible operator for the size of the angular momentum, such as $(\hat{\mathbf{J}}^2)^{1/2}$, is inconvenient. The operator $\hat{\mathbf{J}}$ is also Hermitian, meaning that its three components $\hat{J}_x$, $\hat{J}_y$, $\hat{J}_z$ are all Hermitian.

2. The (length)^2 operator $\hat{\mathbf{J}}^2$ commutes with each of the component operators $\hat{J}_x$, $\hat{J}_y$, $\hat{J}_z$, but the latter do not commute with each other.

3. The eigenvalues of $\hat{\mathbf{J}}^2$ are given by $J(J+1)\hbar^2$ where J is either an integer, or a half-integer and $J \geq 0$. (N.B. do not confuse the operator $\hat{\mathbf{J}}$ with the quantum number J). Sometimes we speak somewhat loosely of J as an 'eigenvalue' or as 'the angular momentum' but strictly it is $J(J+1)\hbar^2$ which is the eigenvalue related to the quantum number J.

4. The eigenvalues of $\hat{J}_z$ are $M_J\hbar$, where M_J has the possible values

$$M_J = -J, -J+1, -J+2, \ldots, J-2, J-1, J \tag{37}$$

Note that this means

$$J \text{ is an integer} \quad \Rightarrow \quad M_J \text{ is an integer}$$
$$J \text{ is a half-integer} \quad \Rightarrow \quad M_J \text{ is a half-integer}$$

5. For a given value of J, there are $2J+1$ possible values of M_J, and therefore $2J+1$ possible different states having angular momentum J. (This is immediately implied by relation (37)).

6. Electrons have spin half. That is, the quantum number associated with their total spin is $s = 1/2$ (it is conventional to use the letter s rather than j for spin angular momentum). Protons, neutrons, muons and various other particles also have spin half.

7. Photons have spin 1. The spin of zero-rest-mass particles is something of a special case, however. It results, for example, in just two (not three) orthogonal polarization states for photons. However the total intrinsic angular momentum remains 1 unit of $\hbar$.

So much for a single angular momentum. Now we will consider what happens when two angular momenta are added together.

Rules for angular momentum: (B) adding angular momenta

8. If a system has two contributions $\hat{\mathbf{j}}_1$ and $\hat{\mathbf{j}}_1$ to its angular momentum, then the operator for the total angular momentum is

$$\hat{\mathbf{J}} = \hat{\mathbf{j}}_1 + \hat{\mathbf{j}}_2$$

9. $\hat{\mathbf{J}}^2$ has eigenvalues $J(J+1)\hbar^2$ where the possible values of J are

$$J = j_1+j_2, \; j_1+j_2-1, \; j_1+j_2-2, \; \ldots \;, |j_1-j_2|+1, \; |j_1-j_2| \tag{38}$$

This makes sense since the angular momenta add like vectors: if the vectors are lined up, the two rotations are going around in the same direction, so the total angular momentum is the sum. If the vectors are pointing in opposite directions, the two rotations are opposed, so the total angular momentum is equal to the larger minus the smaller. If the vectors are at some other angle to one another then the length of the total is in between these limits. The only remarkable thing is that the intermediate values are separated by integers not half-integers. This is related to the fact that the values of M_J are separated by integers.

10. The possible values of J are either all integer, or all half-integer, never some of each. This follows immediately from (38).

Example 1: a single electron.
The total angular momentum operator is $\hat{\mathbf{j}} = \hat{\mathbf{l}} + \hat{\mathbf{s}}$ (just add the operators, rule 6). The spin $s = 1/2$ always, while the orbital angular momentum can be $l = 0, 1, 2, \ldots$. Therefore the total angular momentum can take on the following values:

$$\begin{array}{cccccccc} l = 0 & , & 1 & , & 2 & , & 3 & , & \ldots \\ j = \frac{1}{2} & , & \frac{1}{2} \text{ or } \frac{3}{2} & , & \frac{3}{2} \text{ or } \frac{5}{2} & , & \frac{5}{2} \text{ or } \frac{7}{2} & , & \ldots \end{array} \tag{39}$$

Example 2: a pair of electrons.
We can add the angular momenta up in various ways: We shall see later that the right choice

depends on the relative strengths of the interactions in the system. In the first case, we consider first the total angular momentum of each electron, and then add those. In the second case we consider first the total orbital and total spin angular momentum of the pair, and then add those. Note that while a single electron can only have one value of spin ($s = 1/2$), a pair can have total spin zero or one. It is also possible to add via all the other permutations, such as $[(\hat{\mathbf{l}}_1 + \hat{\mathbf{s}}_2) + \hat{\mathbf{l}}_2] + \hat{\mathbf{s}}_1$ and so on. However, we will see later that in atoms the main two cases are the ones illustrated, especially case (b).

The addition of two spin-half's is very common, and is worth looking into more closely. Consider first the operators. The operator for the total spin of the pair is

$$\hat{\mathbf{S}} = \hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2. \tag{40}$$

Before we consider the eigenstates, it is valuable to ask which of the set of relevant operators commute with one another. The relevant operators are $\hat{\mathbf{S}}$, $\hat{\mathbf{S}}^2$, $\hat{S}_z$, $\hat{\mathbf{s}}_1$, $\hat{\mathbf{s}}_1^2$, $\hat{\mathbf{s}}_{1z}$, $\hat{\mathbf{s}}_2$, $\hat{\mathbf{s}}_2^2$, $\hat{\mathbf{s}}_{2z}$. These two commutation results are standard properties of angular momentum:

$$\left[\hat{\mathbf{S}}^2, \hat{S}_z\right] = 0 \tag{41}$$

$$\left[\hat{S}_x, \hat{S}_z\right] \neq 0. \tag{42}$$

Next, note that

$$[\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2] = 0, \quad [\hat{\mathbf{s}}_1^2, \hat{\mathbf{s}}_2^2] = 0, \tag{43}$$

this is because $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ act on different particles, they do not operate on the same degree of freedom, and therefore must commute. Using this, it is easy to prove

$$[\hat{\mathbf{S}}, \hat{\mathbf{s}}_1] = [\hat{\mathbf{s}}_1 + \hat{\mathbf{s}}_2, \hat{\mathbf{s}}_1] = [\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_1] + [\hat{\mathbf{s}}_2, \hat{\mathbf{s}}_1] = 0 \tag{44}$$

and

$$[\hat{S}_z, \hat{s}_{1z}] = [\hat{s}_{1z}, \hat{s}_{1z}] + [\hat{s}_{2z}, \hat{s}_{1z}] = 0. \tag{45}$$

For each result involving $\hat{\mathbf{s}}_1$, there is a corresponding result involving $\hat{\mathbf{s}}_2$.

The next one takes some work to prove:

$$[\hat{\mathbf{S}}^2, \hat{\mathbf{s}}_1^2] = [\hat{\mathbf{s}}_1^2 + \hat{\mathbf{s}}_2^2 + 2\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2, \hat{\mathbf{s}}_1^2] = 0. \tag{46}$$

The awkward part of the proof is to show $[\hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2, \hat{\mathbf{s}}_1^2] = 0$, it can be done by expanding the dot products into x, y, z components and working through the commutators using the known commutation relations among the components.

Finally, we have

$$[\hat{\mathbf{S}}^2, \hat{s}_{1z}] \neq 0, \tag{47}$$

which is not hard to prove.

Notice the implications of the above. Since the squared operators commute among themselves (eq. (43) and (46)), the states of definite total spin can also be states of definite individual spins, where we mean the size (not direction) of spin. However, because the z-component of spin of either one of the individual particles does not commute with the total $\hat{\mathbf{S}}^2$ (eq. (47)), a state of well-defined total spin will not always have well-defined values for the z-components of the individual spins.

A useful way to picture this is to imagine each spin as a vector, with the individual spin vectors precessing around the total spin which is given by their vector sum (see diagram in the lecture).

The eigenstates of the single-particle spin operators $\hat{\mathbf{s}}^2$, $\hat{s}_z$ are written

$$|s = 1/2,\ m_s = +1/2\rangle \quad = \quad |\uparrow\rangle \tag{48}$$
$$|s = 1/2,\ m_s = -1/2\rangle \quad = \quad |\downarrow\rangle\,. \tag{49}$$

The eigenstates of the total spin and its z component can be written in terms of these eigenstates of the individual spins. The result is

$$|S = 0\rangle \quad = \quad \frac{1}{\sqrt{2}} \left(|\uparrow\rangle_1 |\downarrow\rangle_2 - |\downarrow\rangle_1 |\uparrow\rangle_2 \right) \tag{50}$$

$$\tag{51}$$

$$|S = 1, M_S = 1\rangle \quad = \quad |\uparrow\rangle_1 |\uparrow\rangle_2$$
$$|S = 1, M_S = 0\rangle \quad = \quad \frac{1}{\sqrt{2}} \left(|\uparrow\rangle_1 |\downarrow\rangle_2 + |\downarrow\rangle_1 |\uparrow\rangle_2 \right)$$
$$|S = 1, M_S = -1\rangle \quad = \quad |\downarrow\rangle_1 |\downarrow\rangle_2 \tag{52}$$

The proof of this is given in quantum mechanics text books, it involves use of ladder operators. The main point here is to notice that these results make sense: when the individual spins are both spin-up, the total spin has z component $1/2 + 1/2 = 1$, when one spin is up and the other down, the total spin has z component zero, etc. Also, the states of well-defined total S and total M are not all of well-defined m_1, m_2, this is consistent with the non-commutation result (47).

Note that the there is one state of spin 0, and three states of spin 1, in agreement with rule 5. For this reason the spin 0 state is called a 'singlet' and the three spin 1 states are called a 'triplet'. Note also that the singlet state changes sign if we swap the particles, while the three triplet states are all unaffected by exchanging the particles.

5.1   Main points

The main points of this chapter have been:
meaning and possible values of quantum numbers j, m_j; adding angular momentum operators to make total angular momentum; the possible values of J; the total spin states when adding two spin-half's.

6 Helium

The Schrödinger equation for helium is:

$$H \psi(r_1, \theta_1, \phi_1; r_2, \theta_2, \phi_2) = E \psi(r_1, \theta_1, \phi_1; r_2, \theta_2, \phi_2)$$

where the Hamiltonian is

$$H = \left(\frac{\hat{\mathbf{p}}_1^2}{2m} + \frac{\hat{\mathbf{p}}_2^2}{2m} - \frac{Ze^2}{4\pi\epsilon_0 r_1} - \frac{Ze^2}{4\pi\epsilon_0 r_2} \right) + \frac{e^2}{4\pi\epsilon_0 r_{12}} \tag{53}$$

(I am now expressing the kinetic energy operator for any one particle as $\hat{\mathbf{p}}^2/2m$, which is more convenient to write down). Recall that the quantity $|\psi|^2 dx_1 dy_1 dz_1 dx_2 dy_2 dz_2$ is the probability to find the first electron between x_1, y_1, z_1 and $x_1 + dx_1, y_1 + dy_1, z_1 + dz_1$ *and* the second electron between x_2, y_2, z_2 and $x_2 + dx_2, y_2 + dy_2, z_2 + dz_2$. We can think of helium as the two electrons moving around the stationary nucleus, though strictly speaking it is the motion in the centre of mass frame which we are discussing.

6.1 Main features of the structure

Unfortunately, there is no known way to find an exact solution of this Schrödinger equation by analytical methods. This is the quantum equivalent of the famous 3-body problem which can't be exactly solved in classical mechanics. However, we can get a good understanding by learning from hydrogen and using some common sense.

It turns out that the higher-lying energy levels of helium are quite close to those of hydrogen (see problems). We can understand this as follows: imagine building a helium atom by starting with the bare nucleus, and adding the two electrons. Clearly, when we add the first electron we have a hydrogen-like system with $Z = 2$. Let this first electron occupy the ground state, then it has a spherically symmetric wavefunction. When the second electron is brought in, while it is far from the nucleus it experiences the electric field due to a spherically symmetric charge distribution of both nucleus and inner electron together, making a total charge of 1. This is again a hydrogenic problem, as long as the second electron is far from the nucleus. We take the zero of energy to be the energy of the system when the more distant electron is just able to escape to infinity (ionisation), leaving the other electron in the ground state of the remaining ion. Putting this all together we find the high-lying energy levels will equal those of hydrogen, and we can write

$$\psi(r_1, \theta_1, \phi_1; r_2, \theta_2, \phi_2) \simeq \phi_{1s, Z=2}(r_1) \, \phi_{nl, Z=1}(r_2, \theta_2, \phi_2) \tag{54}$$

where the wavefunctions ϕ are the hydrogen wavefunctions for the given values of nuclear charge, with quantum numbers n, l. For this reason the energy levels of helium for $n > 1$ are labelled 1s2s, 1s2p, 1s3s, 1s3p, 1s3d etc. In fact the wavefunctions will not be exactly equal to hydrogenic ones, because the outer electron wavefunction extends inside the inner one, so the nuclear charge is not fully screened. Also, the inner electron wavefunction is perturbed and the net charge distribution experienced by either electron is no longer quite spherically symmetric. However, it remains a good approximation to treat it as spherically symmetric to first order, so the principal quantum number n and angular momentum quantum number l are still appropriate labels.

So far we have missed out the lowest lying energy level. The second electron has a lowest energy state with principal quantum number 1, giving the state 1s1s for the whole atom, which is written $1s^2$ for brevity. Note that in this case the approximation of spherical symmetry of the net potential experienced by each electron remains very good, so the quantum numbers n, l remain correct labels for this case also, though the screening is now very different from before so the energy level is not

at all close to the hydrogen value... it is **low** down, at -25 eV, because the nucleus is not well screened. Therefore the spectrum of helium has the following pattern: a bunch of transitions way off in the far ultra-violet, and then all the rest at noticeably longer wavelengths (mid ultra-voilet upwards). The former are transitions between ground and excited states, the latter are transitions between one excited state and another.

Note that in helium we are only concerned with excitation of one electron. That is, we don't have states such as $2s^2$, $2s2p$, etc. The reason is simply that it takes more energy to excite both electrons, and in helium and other light atoms the energy required is more than the ionization energy. That is, an atom with this much energy isn't an atom any more, it is a dissociated ion plus free electron (actually, it rapidly decays and becomes an ion plus free electron). In heavier elements, however, excitation of two or more electrons in bound states is possible.

To estimate the ground state energy, we can use a guessed general form of the ground state wavefunction, and apply the variational method. The experimentally observed result is that 24.580 eV is required to remove one electron. The variational procedure has to be applied with considerable skill in the computer programming to obtain this degree of precision.

Note that the ionization potential is very large in helium, indeed it is the largest of any atom—by the end of the course you should be able to give a cogent argument why all the others are lower.

The degree to which each electron screens the other in the $1s^2$ ground state can be obtained from the observed ionisation energy. We argue that before the electron is removed, the total energy of the 2 bound electrons is $2Z_{\text{eff}}^2 E_{\text{R}}$, and after one electron is removed the remaining electron has binding energy exactly $Z^2 E_{\text{R}} = 4E_{\text{R}}$. Therefore

$$
\begin{aligned}
2Z_{\text{eff}}^2 E_{\text{R}} - 4E_{\text{R}} &= 24.58 \text{ eV} \\
\implies Z_{\text{eff}} &\simeq 1.7
\end{aligned}
$$

For higher lying states $1snl$ the approximation that the 1s electron sees the whole nuclear charge Z, while the other sees the effectively screened charge $Z - 1$, works fairly well, and better for high n, high l, for obvious reasons.

[Problem: helium is smaller (has a smaller mean radius) than hydrogen. Why?]

[Problem (tricky): estimate how far one would expect $2s^2$ to be above $1s^2$ in helium].

6.2   Splitting of singlets from triplets

So far we have taken account of the inter-electron repulsion by modelling it in terms of the concept of screening. This involves approximating the resultant effective electric field experienced by each electron as a spherically symmetric field, no matter what the electrons are doing. We would now like to calculate the influence of the feature we have left out, namely the fact that the electrons do influence one another, so they do not move independently. This will involve some subtle ideas, and in order to keep the argument clear I will quote one result which will be proved in the next section.

The only thing I need to quote is the following: *all the energy eigenstates of the helium Hamiltonian*

(53) have the form

$$\psi_+(\mathbf{r}_1, \mathbf{r}_2) = N \left(f(\mathbf{r}_1)g(\mathbf{r}_2) + g(\mathbf{r}_1)f(\mathbf{r}_2) \right) \tag{55}$$

$$\text{or} \quad \psi_-(\mathbf{r}_1, \mathbf{r}_2) = N \left(f(\mathbf{r}_1)g(\mathbf{r}_2) - g(\mathbf{r}_1)f(\mathbf{r}_2) \right) \tag{56}$$

where the constant N simply takes care of normalization, and the two wavefunctions f, g are either the same (in which case the second form vanishes) or orthogonal. It is not hard to prove this using some slightly (but not much) more advanced quantum mechanics[5], but for now let's just trust the result, and notice that it is very similar (in fact the maths is almost identical—see appendix) to the case of coupled oscillators in classical mechanics (discussed in the lecture).

Let's think about the energies of the states ψ_+ and ψ_-. The Hamiltonian can be divided into two parts: everything except the electron-electron repulsion, plus the electron-electron repulsion. Let us call these

$$H_0 = \left(\frac{\mathbf{p}_1^2}{2m} + \frac{\mathbf{p}_2^2}{2m} - \frac{Ze^2}{4\pi\epsilon_0 r_1} - \frac{Ze^2}{4\pi\epsilon_0 r_2} \right) \tag{57}$$

and $H^r = e^2/4\pi\epsilon_0 r_{12}$. What we know so far is that $(H_0 + H^r)\psi_\pm = E_\pm\psi_\pm$. We will now make the *approximation* that f and g are both eigenstates of H_0. This approximation is not in fact needed in a full argument, but it makes the discussion clearer here, and the main result we find will still be valid[6].

Since f and g are both eigenstates of H_0, it is quite easy to see that $H_0\psi_\pm = (E_f + E_g)\psi_\pm$. In other words, states ψ_+ and ψ_- have the same energy as far as kinetic energy and interactions with the nucleus are concerned. Next, we can see that the repulsion H^r increases the energy of both states but gives a higher energy to the ψ_+ state than the ψ_- state. This is because the parts of wavefunctions f and g which overlap interfere constructively in ψ_+, but cancel each other out in ψ_-. These parts correspond to the two electrons being close together. Therefore the electrons are more likely to be close together in ψ_+ so they have more Coulomb repulsion energy. These statements imply that the motions of the electrons are *correlated* for states such as $\psi_\pm$.

The conclusion is that for each state of helium of given n_1, l_1, n_2, l_2, we expect to find two energy levels, owing to the effect of the electron-electron interaction, except for the case when $n_1 = n_2$ and $l_1 = l_2$ since then ψ_- vanishes. When we examine the helium spectrum, and deduce the energy levels, we find that this is exactly what is observed. Good! However, a further strange feature is also observed. When we look at the total spin of each state, we find we do not get all the possibilities. The energy level lying higher, therefore the ψ_+ state, is found always to have total spin 0, not spin 1; while the energy level lying lower, therefore the ψ_- state, is found always to have spin 1, never 0. **The reason for this is one of the most remarkable and beautiful features of quantum physics**, which I will now elaborate.

6.3 Exchange symmetry and the Pauli exclusion principle

One of the most important, perhaps *the* most important idea in physics is that of *symmetry*. By this we mean not just mirror reflection, but symmetry as the general concept of a physical system staying unchanged when something is done to it. For example, a sphere is symmetric with respect to rotations, and the behaviour of any group of particles doesn't change when we displace the system in space. system, a property referred to as translational symmetry. The latter idea can

---

[5]Just examine the exchange symmetry of the Hamiltonian, i.e. find that $[H, \chi] = 0$ where χ is the exchange operator.

[6]A more accurate treatment could employ the self-consistent field method discussed in section 7.2

be used, through a quantum mechanical argument, to derive that momentum is conserved! (See Feynman Lectures on Physics, vol 1, chapter 52). This is a truly staggering insight. It is also a strong hint at how powerful a concept symmetry is in physics, especially quantum physics.

The concept of symmetry allows us to see, for example, that the electric field produced by a spherically symmetric charge distribution must also be spherically symmetric. It must be, since if there were some non-symmetric kink in the field, *how would the electric field 'know' where the kink should go?* Indeed such a situation would be like a contradiction: if the field produced by some object were not spherically symmetric, if follows that the object can't have been spherically symmetric after all.

One can think of this in terms of information. The laws of the universe are cagey: Nature is like a poker player holding her cards close to her chest, never allowing any physical phenomenon to reveal more information in any situation than the minimum logically required by the conditions. So for example, a kink in an electric field gives information by picking out a direction in space, but if there is no direction picked out by the underlying charge distribution, this information has appeared from nowhere, which does not happen.

In quantum mechanics, the principle of symmetry applies with a vengeance. If there is nothing to decide whether an electron should go through one slit or another, then it will go through both! (this is the Young's slits experiment). Furthermore, if there is nothing to distinguish between two particles such as two electrons or two photons, then *the quantum state of two identical particles must not give any information which depends on which particle is which.* This **exchange principle** can be mathematically expressed as follows:

> **Exchange principle**: If $\psi(\{q\}_1, \{q\}_2)$ is the state of two identical particles which have sets of coordinates[7] (such as position and spin) $\{q\}_1$ and $\{q\}_2$, where 1 and 2 label the particles, then
>
> $$|\psi(\{q\}_1, \{q\}_2)|^2 = |\psi(\{q\}_2, \{q\}_1)|^2 \tag{58}$$

Or, in words, the modulus squared of the wavefunction does not change when the labels on the particles are exchanged. Note that we take the modulus squared of the wavefunction because the physically observable quantities depend on $|\psi|^2$ not on ψ directly. This is important, because it leads to two possibilities, both of which are observed in Nature[8]:

$$\text{either} \quad \psi(\{q\}_1, \{q\}_2) = \psi(\{q\}_2, \{q\}_1) \quad \text{(Bosons)} \tag{59}$$
$$\text{or} \quad \psi(\{q\}_1, \{q\}_2) = -\psi(\{q\}_2, \{q\}_1) \quad \text{(Fermions)} \tag{60}$$

If you discover a relationship like this, or a remarkable implication of it, then you might get half the particles in the universe named after you! The first case is called symmetric with respect to exchange of particles, the second case is called antisymmetric with respect to exchange of particles.

Once a boson always a boson: if any one pair of photons is in a symmetric state, then all pairs of photons must be in symmetric states, since otherwise we could distinguish some photons from others. Similarly for fermions. We find in practice that particles of half-integer spin such as electrons and protons are fermions, while particles of integer spin such as photons are bosons. This

---

[7]Here the word "coordinates" is being used in a generic way, to refer to any complete set of physical parameters which are needed to specify a state, including both spatial and spin "coordinates".

[8]Why only two possibilities ± 1, not an infinite number, corresponding to all values of phase $\exp(i\phi)$? The point is there are at most two *linearly independent* possibilities, and we choose the real numbers ± 1 because exchanging particles is a physically possible operation, so is associated with a Hermitian operator having real eigenvalues. See Leighton or a quantum mechanics textbook for details.

relationship between spin and exchange symmetry is called the spin-statistics theorem, which may itself be derived from a subtle higher-order symmetry which is beyond the scope of this course.

Let's consider some implications of (59) and (60). To keep things simple we will concentrate on the spatial part of the total states (the illustrative examples in this section will assume pairs of particles whose spin state is symmetric under exchange of particle labels). Suppose we have a photon in some state $\phi(\mathbf{r}_1)$. Suppose another photon is then placed in the the same state $\phi(\mathbf{r}_2)$ (this means the second photon overlaps the other, having the same frequency and direction). The total state of the two photons is then

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \phi(\mathbf{r}_1)\,\phi(\mathbf{r}_2) \tag{61}$$

Now, although we called the photons 1 and 2, there is nothing in this state which depends on which photon is called 1, and which is called 2. It is therefore a possible state and it obeys the boson relation (59). Photons from a laser are like this: lots of them in the same quantum state.

Now consider two different states f and g for a photon. For example, these could be states of the same frequency and polarization, but different direction (think of two laser beams shining onto a screen with a small angle between them). If the first photon is in state $f(\mathbf{r}_1)$, then the second photon is in $g(\mathbf{r}_2)$. However, who told us which photon to put in f, and which in g?—the state could equally be $f(\mathbf{r}_2)$ combined with $g(\mathbf{r}_1)$. When there is nothing to choose between two cases, both occur, and the true state of the photons will be

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}\left(f(\mathbf{r}_1)\,g(\mathbf{r}_2) + f(\mathbf{r}_2)\,g(\mathbf{r}_1)\right). \tag{62}$$

Make sure you convince yourself that this state obeys the Bose relationship (59).

Now consider the experiment we are discussing: two laser beams illuminating a screen. The light hitting the screen is composed of two plane waves with a small angle between them. This is like the Michelson interferometer with the mirrors at a small angle, or the Young's slits experiment: an interference pattern of straight lines will be observed on the screen. Now, if each of the photons arrived simply in one beam or the other, then we would never have an interference effect. However, fringes are observed. Think hard about the fact that this is a really remarkable observation: if each photon emerges from *either* one laser *or* the other, which is surely what we would guess, then no interference fringes would be seen. Yet they are seen. In case you are not sufficiently impressed by this observation, perhaps because you think photons are insubstantial, you may be interested to learn that the same experiment can be done, and has been done, with *independently prepared* beams of Bosonic atoms (obtained by allowing two small cold gas samples to fall under gravity until they overlap; the atoms are observed to arrive at a detecting screen only along a series of lines, consistent with interference of their de Broglie waves). But interference only happens when one quantum entity (here, an atom) interferes with *itself*—or that is what we learn from the analysis of single-particle quantum mechanics—so does this mean each atom existed all along in both gas samples?

The symmetric state (62) implies that each photon of any pair of photons is in some sense in both beams at once, even though the beams were produced by independent lasers!

To avoid making dubious statements about 'a photon' being produced in two independent lasers at once, or 'an atom' being prepared in two independent atom traps at once, we have to take seriously the idea that the Bosons are truly indistinguishable, so that in situations like this it is misleading to speak of 'this photon' or 'that photon' as if they were identifiable, and even more remarkably it is misleading to speak of 'this atom' or 'that atom'[9].

---

[9]N.B. it is a very severe test of our physical intuition to arrive at the clearest way to describe this experiment in terms of photons: your tutors will probably not all agree on which statements are least misleading, though they will agree on the fringe pattern observed.

It is instructive to consider also the case of two fermions such as two electrons. This time the total state of is type (60). Suppose two electrons are both in some state ϕ. Then the total state would be $\psi = \phi(\mathbf{r}_1)\phi(\mathbf{r}_2)$, but this has the wrong symmetry so is not allowed according to (60). To understand what is going on, take the case where the two electrons are in different states f and g. Then the total state, with the correct symmetry (i.e. antisymmetric with respect to exchange) is

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} \left(f(\mathbf{r}_1)\, g(\mathbf{r}_2) - f(\mathbf{r}_2)\, g(\mathbf{r}_1) \right). \tag{63}$$

Now, observe what happens if we let g become the same as f: in that case ψ vanishes. The upshot of this is

The Pauli Exclusion Principle: *no two identical Fermions can ever be in the same quantum state.*

The Pauli Exclusion Principle underlies atomic structure through and through, as we will see.

6.4   Exchange symmetry in helium

We can now see the reason why the energy levels of helium had the property we remarked at the end of section 6.2. This was the property that each state of given electrostatic energy does not have every possible value of spin (where we have included the full influence of electron-electron repulsion in the electrostatic energy).

The answer comes by putting together the spin states (50), (52), the spatial states (55), (56), and the exchange antisymmetry for fermions (60).

Since the spin and spatial parts of the total Hamiltonian H are merely added together (indeed, the spin part is zero), the solutions can be written as a *product* of spatial and spin states. Since our spatial solutions (55),(56) have definite symmetry (w.r.t. exchange of the particles), we can ensure an overall antisymmetric state by requiring the spin parts to have definite symmetry also, and in each case the opposite symmetry to the spatial state. Therefore the total state is

either   (symmetric spatial state) $\times$ (antisymmetric spin state)
or   (antisymmetic spatial state) $\times$ (symmetric spin state)

Now, the spin zero or 'singlet' state is antisymmetric, (see equation (50)) while the spin one or 'triplet' states are all symmetric (see (52)), so we find the ψ_+ spatial state (55) is only associated with a spin of zero, while the ψ_- spatial state (56) is only associated with a spin of one. The ψ_+ has more electrostatic repulsion energy than the ψ_- latter, so for a given set of quantum numbers n_1, l_1, n_2, l_2, the singlet state has more energy than the triplet state.

The summary is as follows.
1. States which are degenerate under the central field Hamiltonian are shifted and split by the electron-electron interaction Hamiltonian.
2. The spatial parts of the eigenstates of the total Hamiltonian have definite symmetry w.r.t. exchange of the electrons, the symmetric spatial state having more energy than the antisymmetric one.
3. The requirement of asymmetry for identical fermion states means that each spatial state must be associated with a specific spin state.
4. The combination of (1)-(3) leads to an energy splitting between spin singlet and triplet states, even though no spin-dependent interactions are involved.
5. The singlet terms are the higher in energy since when the spatial state is symmetric, the mean separation of the electrons is smaller, leading to a higher energy of repulsion between them.

One can readily see that the above summary will carry over to the case of many electrons. This explains the splitting between singlet and triplet states observed in all atoms from group 2 of the periodic table, and related observations in all atoms of more than one electron.

6.5 Main points

The main points of this chapter have been:
Schrodinger equation for helium; concept of screening of nuclear charge (appealing to Gauss' theorem of electrostatics); deduction that all energy levels other than ground state are close to hydrogen values; the ground state energy is very much lower since screening is especially imperfect when both electrons are in the same spatial state; symmetry of the wavefunctions with respect to exchange of the particle labels; Exchange Principle; Bosons and Fermions; Pauli Exclusion Principle; deduction that in Helium spatial states of given exchange symmetry must be associated with spin states of the opposite exchange symmetry, and hence that only one spin state is associated with each spatial energy eigenstate; fact that the spatial states have energy contribution from the electron-electron repulsion, and hence the overall implication that each triplet lies lower in energy that the corresponding singlet.

6.5.1 Appendix: detailed derivation of states and splitting

I will do the calculation for helium, starting by neglecting the electron-electron repulsion all together to begin with. This serves to bring out the features of the calculation, which we will then make more precise at the end. The level of this material is just about on the syllabus: you would not be expected to reproduce the full derivation in detail, but you should be familiar with the main points.

So, considering just the attraction to the nucleus, we have the Hamiltonian

$$H_0 = \left(\frac{\mathbf{p}_1^2}{2m} + \frac{\mathbf{p}_2^2}{2m} - \frac{Ze^2}{4\pi\epsilon_0 r_1} - \frac{Ze^2}{4\pi\epsilon_0 r_2} \right) \tag{64}$$

where $Z = 2$. Let f and g be eigenstates of the single-electron Hamiltonian $\mathbf{p}^2/2m - Ze^2/4\pi\epsilon_0 r$, with eigenvalues E_a and E_b respectively. Then, by the fact that H_0 is separable, the product states

$$u_{fg} \equiv f(\mathbf{r}_1)g(\mathbf{r}_2),$$
$$u_{gf} \equiv g(\mathbf{r}_1)f(\mathbf{r}_2),$$

are eigenstates of H_0, both having the eigenvalue $E_0 = E_f + E_g$. Note that the separation of variables method has converted our single Schrödinger equation into the pair of equations

$$\left. \begin{array}{rcl} H_0 u_{fg} &=& E_0 u_{fg} \\ H_0 u_{gf} &=& E_0 u_{gf} \end{array} \right\} \tag{65}$$

If we write a general combination of u_{fg} and u_{gf} as $\alpha u_{fg} + \beta u_{gf}$ where α and β are any complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$, then the pair of Schrödinger equations can be expressed as a single matrix equation

$$\begin{pmatrix} \langle fg| H_0 |fg \rangle & \langle fg| H_0 |gf \rangle \\ \langle gf| H_0 |fg \rangle & \langle gf| H_0 |gf \rangle \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = E_0 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \tag{66}$$

which is just the simple expression

$$\begin{pmatrix} E_0 & 0 \\ 0 & E_0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = E_0 \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{67}$$

We have a degeneracy (two states of the same energy), which will be removed by the electron-electron interaction $H^r = e^2/4\pi\epsilon_0 r_{12}$. Let's assume that this interaction is small enough that the energy shifts of the initially degenerate states u_{fg} and u_{gf} are primarily due to the coupling between them, rather than to the influence of further states of energy $\neq (E_a + E_b)$. Hence we can approximate the problem as a two-state system, with the Schrödinger equation

$$\begin{pmatrix} E_0 + H_{11}^r & H_{12}^r \\ H_{21}^r & E_0 + H_{22}^r \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = E \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

where

$$
\begin{aligned}
H_{11}^r &= \langle fg| H^r |fg\rangle = \int f^*(\mathbf{r}_1)g^*(\mathbf{r}_2)H^r f(\mathbf{r}_1)g(\mathbf{r}_2)dV_1 dV_2 \\
&\equiv J \quad [\text{`direct' integral} \\
H_{22}^r &= \langle gf| H^r |gf\rangle = \int g^*(\mathbf{r}_1)f^*(\mathbf{r}_2)H^r g(\mathbf{r}_1)f(\mathbf{r}_2)dV_1 dV_2 \\
&= H_{11}^r \quad [\text{ since } H^r \text{ only depends on relative coordinate } r_{12} \\
H_{12}^r &= \langle fg| H^r |gf\rangle = \int f^*(\mathbf{r}_1)g^*(\mathbf{r}_2)H^r g(\mathbf{r}_1)f(\mathbf{r}_2)dV_1 dV_2 \\
&\equiv K \quad [\text{`exchange' integral} \\
&= H_{21}^r \quad [\text{real and Hermitian } H^r
\end{aligned}
$$

The names 'direct' and 'exchange' integrals come from the fact that the direct integral can be interpreted as the mutual energy caused by repulsion between two overlapping charge distributions $|f|^2$ and $|g|^2$, while the exchange integral does not have such a simple interpretation. This is simply an example of the problems which can arise when trying to interpret a two-particle quantum state in terms of single-particle states. It has become common practice to call the off-diagonal part of H^r an 'exchange interaction', but this is to some extent misleading because there is no new interaction here: it is just the *electrostatic interaction* acting on a two-particle state.

Our Schrödinger equation now has the very simple form

$$\begin{pmatrix} E_0 + J & K \\ K & E_0 + J \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = E \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

which is just that of the familiar situation of coupled pendula. By now we know how to solve this in our sleep, $(E_0 + J - \lambda)^2 - K^2 = 0 \Rightarrow E_0 + J - \lambda = \pm K$, and the answer is:
eigenvalues: $E_0 + J + K$, $E_0 + J - K$
eigenstates: $\beta = \alpha$, $\beta = -\alpha$.

Therefore the two-electron energy eigenstates are $u_{fg} \pm u_{gf}$, i.e.

$$\frac{f(1)g(2) + g(1)f(2)}{\sqrt{2}}, \quad \frac{f(1)g(2) - g(1)f(2)}{\sqrt{2}}, \tag{68}$$

and the first of these has more energy than the second. The physical reason why the first state has more energy is that in this symmetric state, those parts of f and g which overlap interfere constructively, making an enhanced probability for the electrons to overlap (and hence have a higher

energy of interaction) while in the antisymmetric state the overlapping parts interfere destructively, so the electrons are further apart and have less energy of interaction.

In the case of coupled classical pendulums we would have similar results: one normal mode has a higher energy than the other because the different *correlations* in the motions of the pendulums cause a different energy from the coupling.

Note that we have obtained the answer *without appealing either to the spin or to the exchange symmetry of the states*. As far as spin goes, at the moment it looks as though each spatial wavefunction can be associated with any spin state. As far as symmetry goes, the electrons could be distinguishable, for all we care (e.g. one painted blue, somehow, and the other painted red). N.B. many (in fact the majority) of textbooks are quite confused on this point, or downright wrong.

Now, as it happens, we are dealing with two identical electrons, rather than one 'red' and one 'blue'. Therefore, by the principles of quantum mechanics as they apply to identical particles, the total state must be one of *definite symmetry* with respect to exchange of the particles. Indeed, since electrons are fermions the state must be antisymmetric. Therefore at first sight only the second of our two solutions (68) looks physically possible. In fact both solutions are possible because each electron has a further degree of freedom, namely its spin, but the implication is that not every spin state can go with every spatial state, as discussed in section 6.4.

An exact (within numerical error) treatment would involve a full (numerical) solution of the Schrödinger equation including all the parts of the Hamiltonian from the start. The solutions are then found to have the same general features as given here, but the precise values of energy and the shape of the wavefunctions would differ from those given by the approximate method.

7 Many electrons, a first look

Hydrogen and helium have now taught us most of the concepts we need to understand any atom. What remains is a more careful understanding of the idea of screening, a more thorough discussion of the angular momentum of many electrons, and various effects which are small in size but large in physical significance; they are treated using using perturbation theory.

7.1 The periodic table

Here is the Hamiltonian for an atom containing many electrons:

$$H = \sum_i \left(\frac{\mathbf{p}_i^2}{2m} - \frac{Ze^2}{4\pi\epsilon_0 r_i} \right) + \sum_{i>j} \frac{e^2}{4\pi\epsilon_0 r_{ij}} + \dots \tag{69}$$

Where the $\dots$ indicates that we have missed out the small effects which will be considered later on (mainly spin-orbit coupling and hyperfine structure). The symbol $\sum_{i>j}$ means we sum over all values of $j = 1$ to N (for N electrons) and values of i from $j+1$ to N (thus avoiding counting pairs twice).

The numerical problem of solving Schrödinger's equation with more than a few electrons is extremely difficult. It is necessary to simplify the task, and the *self-consistent field*, or Hartree method (and further developments based on it) is the best way to do this. The idea is to reason that each electron moves in the total electric field produced by the nucleus and all the other electrons, and this total electric field is close to spherically symmetric (though with a complicated dependence on r). Therefore we can approximate the true Hamiltonian with a spherically symmetric one. The one we try is at first based on a reasonable guess, and then the guess is improved by an iterative procedure which will be presented in the next section.

The total spatial wavefunction of the atom is, in this approximation, a product of single-electron wavefunctions, where each wavefunction is of the type which arises in a central field, and therefore has an angular part given by $Y_{l,m}(\theta, \phi)$ and a radial part which depends on the two quantum numbers n and l. This means that we can specify the total wavefunction simply by listing the quantum numbers n and l for all the electrons (we don't bother to mention m because the energy does not depend on it, but we keep in mind that a given l has $2l+1$ values of m associated with it). The list of quantum numbers is called the *configuration* of the atom. We have in fact already met some configurations of helium, such as 1s^2, 1s2s, 1s2p, 1s3s etc. [Problem: why not 1s1p or 1s2d?]

Actually, in order to obey the exchange symmetry rules, the true total spatial wavefunction is either an antisymmetric or a symmetric combination of the ones indicated by the configuration, but as long as we take this for granted then we can carry on referring to 'an electron in the 3p state' etc.

Our next element is lithium ($Z = 3$). The ground state is therefore 1s1s1s, or 1s^3 for short??? No! This state is ruled out by the Pauli Exclusion Principle. We can have at most two electrons in any given spatial state—the two electrons then must have different spin states, which is possible since a spin-half particle has two different possible values for m_s, namely $+1/2$ and $-1/2$. Therefore there can be at most 2 electrons in a state having $l = 0$. For $l = 1$ we can have 6 electrons however, since there are three possible values of m_l, and for each m_l there are still the two possible spin states. Thus the general rule is that an l state can contain $2 \times (2l+1) = 4l+2$ electrons.

The ground state configurations of the elements can now be written down, and suddenly all the

basic principles of chemistry spring into view! For, going from each element to the next we obtain the following ground state configurations:

element	atomic number Z	ground state configuration
H	1	1s
He	2	$1s^2$
Li	3	$1s^2\,2s$
Be	4	$1s^2\,2s^2$
B	5	$1s^2\,2s^2\,2p$
C	6	$1s^2\,2s^2\,2p^2$
N	7	$1s^2\,2s^2\,2p^3$
O	8	$1s^2\,2s^2\,2p^4$
F	9	$1s^2\,2s^2\,2p^5$
Ne	10	$1s^2\,2s^2\,2p^6$
Na	11	$1s^2\,2s^2\,2p^6\quad 3s$
etc.		

I have drawn a horizontal line where the principal quantum number n of the last electron is forced by Pauli exclusion to go up 1. The states of given n are referred to collectively as a *shell*, since the electrons sit or orbit further from the nucleus as n increases. States of given n and l form a subshell. I won't write a discussion of the periodic table here, since that is done perfectly well in all the textbooks, and is quite straightforward. There are many interesting subtleties, the most prominent of which is to notice that when we get to $Z = 19$ (potassium)

the configuration is $\quad 1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,$ **4s**

when we might have expected $\quad 1s^2\,2s^2\,2p^6\,3s^2\,3p^6\,$ **3d**

I have highlighted the last electron for clarity. The point is that the 4s state lies lower in energy than the 3d state for potassium. This is because the $l = 0$ states penetrate in towards $r = 0$ more than the states of higher l, and therefore the $l = 0$ electrons experience less screening of the nuclear charge, which is now quite large. This lowers the energy, and in this case the lower potential energy of 4s more than compensates the increased kinetic energy associated with its larger value of n, compared to 3d. The transition metals fill in the region where the $4s^2$ states are both filled, so the 3d states are next filled ... but have a look at the exact ordering and you will see some departures from this simple rule.

In a given configuration, the filled shells, also called 'closed' shells, form the *core* and the electrons lying outside the core are called *valence* electrons.

Note that since

$$\sum_{m=-l}^{l} |Y_{l,m}|^2 \text{ is independent of } \theta, \phi$$

the core is always spherically symmetric.

7.2 The self-consistent field method

Now we will examine the way the central field is calculated. As previously noted, the idea is to guess a central field which reasonably approximates the total effect of nucleus and electrons, and then the guess is improved by an iterative procedure, as follows:

1. Set $H = \sum_i \left(\mathbf{p}_i^2/2m + V(r_i) \right)$ where $V(r_i)$ is based on a reasonable guess. Note that here all the electrons experience the same field, and the equation is a simple sum of terms for different

electrons, therefore it has solutions in the form of a product of wavefunctions of different electrons. In physical terms, we picture all the electrons as moving *independently* in the net field. This is quite a surprising physical picture, but it makes good sense mathematically.

2. Numerically solve the single-electron Schrödinger equation with $H = \mathbf{p}_i^2/2m + V(r)$ (easy on a computer), obtaining all the eigenvalues and eigenstates $\phi_{n,l}$.

3. Assume the many-electron wavefunction is a product of the N lowest energy wavefunctions (if we are trying the calculate the ground state), with two electrons per wavefunction, as in the configurations already discussed.

4. Calculate the charge distribution $-e|\phi_{n,l}|^2$ of each electron, and add them up to get the net charge distribution. Then work out from standard electrostatics what the potential energy of an electron would be in the combined field of this charge distribution and the nucleus (see comment below). Let $V'(r)$ be this potential energy.

5. $V'(r)$ is now an improved guess for the correct central field, so if it isn't already equal to the $V(r)$ we used last time, then return to step 2, now using $V'(r)$.

6. If $V'(r) = V(r)$ (to within some very small difference) then we have a self-consistent field. That is, the field we assumed is the very one which the electrons and nucleus are producing.

There is a subtlety which arises in step 4. The field which any given electron experiences has to be not the total field of all Z electrons plus nucleus, but the net field produced by the $Z - 1$ *other* electrons plus the nucleus. On the other hand, the approximation we are making is that all electrons experience the same net field. One way to get a self-consistent argument is to take the net charge distribution of all Z electrons and multiply it by $(Z - 1)/Z$ to obtain the charge distribution whose field is experienced by any one electron.

The method as described is the Hartree method. The Hartree-Fock method is the same, except that a fully antisymmetric form for the wavefunctions is used, which makes the calculations harder and the results more accurate. Further improvements are to take account of relativity, and so on.

7.3 Group 1: the alkalis; quantum defect

Atoms in group 1, the alkalis, have just one electron outside closed subshells. Therefore the ground state has a configuration of the form $\cdots ns$ where $n = 1$ for hydrogen, 2 for lithium, 3 for sodium, and so on.

The excited states of a given atom have configurations consisting of a fixed core, with the last electron excited to higher levels such as 3p, 3d, 4s, etc. Therefore the energy level diagram is labelled by just the quantum numbers of this last electron. The simple idea of screening which we met in helium is quite a good approximation because the core is spherically symmetric. The valence electron sees the nuclear charge Z screened by $Z - 1$ electron charges in a spherical core, so the energy levels are given approximately by the hydrogen formula $-E_R/n^2$ (where $E_R = (1/2)m\alpha^2 c^2$). To make the formula more accurate, we could introduce an effective nuclear charge slightly larger than 1 to account for the fact that the screening is never perfect, but it is more useful to introduce a term δ called the quantum defect, which adjusts n in the formula:

$$E_n = -\frac{E_R}{(n - \delta(l))^2} \tag{70}$$

The reason δ is useful is that it turns out that when we fit all the observed energy levels to this formula, the value of δ depends mostly on l and very little on n. This useful behaviour would not

happen if we used an effective nuclear charge instead, since that would clearly have to tend to 1 as n or l increased. δ is larger for s states ($l = 0$) because they penetrate the core more. Note that since δ does not vary much with n, in the limit of large n the formula tends to the hydrogen one, which makes sense since then the valence electron is lying far outside the core so the screening is very good.

Quantum defect theory enables us to predict the positions of further levels from information about a few: see the problems. Here are two simple observations which you should be able to make sense of quite quickly: (1) all the alkalis have similar ionization potentials (i.e. within a factor 1.5 even though the nuclear charge varies by a factor 29); (2) the principle transition (i.e. from ground state to nearest p state) in the alkalis is always in or near the visible region of the spectrum.

7.4 Group 2: alkaline earths

In group two the atoms have 2 valence electrons. The main new feature is that the low-lying configurations give two energies rather than one in the gross structure, because as with helium the two valence electrons can be in an antisymmetric or a symmetric state. A more full discussion is postponed until section 11 after we have dealt with the next most important terms in the Hamiltonian. The excited states can involve excitation of one or both valence electrons (thus for example 3s3p, 3s4s, 3s3d, ..., $3p^2$, 3p4s, ... $4s^2$, 4s4p, ... for magnesium), but to excite one electron requires less energy than to excite two, so the low-lying states have configurations of the form $nsn'l'$ where n is the principal quantum number of the ground state, and $n' \geq n$.

7.5 Fun with screening

It is possible with a bit of common sense to make a lot of simple observations on the basis of screening. We already noted, for example, that the quantum defect in alkalis will get smaller as l increases because then the valence electron wavefunction lies more and more outside the core. A further general observation is that an ion will be closer to hydrogenic form than the neutral atom having the same number of electrons. This is because the nuclear charge is larger so dominates the electronic charge better. This is why the quantum defects of levels in Al^{++} ($Z = 13$) are smaller than the defects of the corresponding levels in Na ($Z = 11$) for example. In K ($Z = 19$) the 3d configuration lies above 4s and 4p and even above 5s, while in the calcium ion Ca^+ ($Z = 20$) the 3d configuration lies between 4s and 4p, which is closer to hydrogenic ordering.

7.6 Where we are heading

We now have a good understanding of the atom in the central field approximation. In the next chapter we will apply this to the X-ray spectra of atoms, where we will see that the central field approximation works very well, though not perfectly of course.

After that, the next step is to take account of the part of the electric field in the real atom which is not included in the central field, and to take account of further effects which are also small compared to those considered up till now. To do all these things we will use perturbation theory.

7.7   Main points

The main points of this chapter have been:
The central field approximation and configurations; the Periodic Table; spherical symmetry of the core; self-consistent method to calculate appropriate potential energy function; quantum defect δ for alkalis, mainly a function of l; screening arguments.

8 X rays

For this subject I refer you to Dr Ewart's notes, and the textbooks. Kuhn is thorough on X rays, and up-to-date comments are in Haken and Wolf. Eisberg and Resnick is also good.

9 Degenerate Perturbation Theory

You have met perturbation theory in the quantum mechanics lectures. The main result is as follows. If a system has a Hamiltonian of the form

$$H = H^0 + \delta H \tag{71}$$

where the effects of δH are small compared with those of H^0, then we can solve the Schrödinger equation

$$H\psi_k = E_k \psi_k \tag{72}$$

by using the form

$$\psi_k = \psi_k^0 + \delta\psi_k$$
$$E_k = E_k^0 + \delta E_k$$

where ψ_k^0 and E_k^0 are the eigenfunctions and eigenvalues of H_0, i.e.

$$H^0 \psi_k^0 = E_k^0 \psi_k^0 \tag{73}$$

The notation has the following logic. We use the subscript k to count the different eigenstates, and the superscript zero is applied to quantities related to the zeroth order Hamiltonian H^0 which is present before the perturbation is applied.

Plugging all this into (72) we have

$$\left(H^0 + \delta H\right)\left(\psi_k^0 + \delta\psi_k\right) = (E_k^0 + \delta E_k)\left(\psi_k^0 + \delta\psi_k\right)$$

Multiplying out the brackets we obtain:

$$H^0\psi_k^0 + H^0\delta\psi_k + \delta H\psi_k^0 + \delta H\delta\psi_k = E_k^0\,\psi_k^0 + E_k^0\,\delta\psi_k + \delta E_k\,\psi_k^0 + \delta E_k\,\delta\psi_k \tag{74}$$

and then ignoring 2nd order small quantities we have

$$H^0\psi_k^0 + H^0\delta\psi_k + \delta H\psi_k^0 \simeq E_k^0\,\psi_k^0 + E_k^0\,\delta\psi_k + \delta E_k\,\psi_k^0$$
$$\Rightarrow H^0\delta\psi_k + \delta H\psi_k^0 = E_k^0\,\delta\psi_k + \delta E_k/, \psi_k^0 \tag{75}$$

Where we used (73). Now multiply by $(\psi_k^0)^*$ and integrate:

$$\Rightarrow \left\langle\psi_k^0\right|H^0\left|\delta\psi_k\right\rangle + \left\langle\psi_k^0\right|\delta H\left|\psi_k^0\right\rangle = \left\langle\psi_k^0\right|E_k^0\left|\delta\psi_k\right\rangle + \left\langle\psi_k^0\right|\delta E_k\left|\psi_k^0\right\rangle$$
$$\Rightarrow \left\langle\psi_k^0\right|\delta H\left|\psi_k^0\right\rangle = \delta E_k \tag{76}$$

The last step used the result $\left\langle\psi_k^0\right|H^0\left|\delta\psi_k\right\rangle = E_k^0\left\langle\psi_k^0\right|\delta\psi_k\rangle$. To prove this, use the fact that H^0 is Hermitian, which implies that $\left\langle\psi_k^0\right|H^0\left|\delta\psi_k\right\rangle = \left\langle\delta\psi_k\right|H^0\left|\psi_k^0\right\rangle^*$, and then examine the integral, use (73) and to finish recall that the energy E_k^0 is real. The more expert quantum mechanic can see it immediately by operating H^0 to the left inside the bracket, but if you don't know what I mean by that, then ignore it.

So far we have derived the important result that a small perturbing Hamiltonian will shift the energy of the k'th state by an amount equal to the expectation value of the perturbation in the original state:

$$\delta E = \left\langle\psi_k^0\right|\delta H\left|\psi_k^0\right\rangle. \tag{77}$$

Next we would like to know the effect on the wavefunction. To find this, return to equation (75) and multiply by one of the eigenstates different from ψ_k^0 before integrating:

$$\langle \psi_n^0 | H_0 | \delta\psi_k \rangle + \langle \psi_n^0 | \delta H | \psi_k^0 \rangle = \langle \psi_n^0 | E_k^0 | \delta\psi_k \rangle + \langle \psi_n^0 | \delta E | \psi_k^0 \rangle \tag{78}$$

$$\Rightarrow \quad \langle \psi_n^0 | \delta H | \psi_k^0 \rangle = \langle \psi_n^0 | \delta\psi_k \rangle (E_k^0 - E_n^0) \tag{79}$$

using the same reasoning as before. Now, the change in wavefunction $\delta\psi_k$ can always be written in terms of the zeroth order wavefunctions, since they form a complete set:

$$\delta\psi_k = \sum_n c_n \psi_n^0 \qquad \text{where } c_n = \langle \psi_n^0 | \delta\psi_k \rangle.$$

The perturbation theory result (79) tells us the coefficients, so we have

$$\delta\psi_k = \sum_{n \neq k} \frac{\langle \psi_n^0 | \delta H | \psi_k^0 \rangle}{E_k^0 - E_n^0} \psi_n^0. \tag{80}$$

The term $n = k$ is omitted from this sum: this is ok since any contribution from ψ_k^0 to $\delta\psi_k$ is taken care of at the end by imposing an overall normalisation of the new state $\psi_k^0 + \delta\psi_k$ (to first order approximation).

As long as for the energy level k we are considering, no other level has the same energy, then all is well. However, if there is degeneracy $E_n^0 = E_k^0$ then the answer blows up. The rest of this chapter is dedicated to a careful consideration of how to handle this very common situation.

9.1 An exact analysis of a simple example

To get some insight into perturbation theory in the degenerate case, we will consider an example where we can work out the exact answer without using perturbation theory. The example we will examine is a 2-dimensional harmonic potential well $V_0 = (1/2)m\omega^2(x^2 + y^2)$, with a perturbation to be described in a moment.

The situation before the perturbation is applied can be analyzed by separation of variables, giving solutions of the form $\psi_0(x, y) = u_0(x)v_0(y)$. The energy levels of the 1-dimensional Schrödinger equations are $E_{x,m} = (m + 1/2)\hbar\omega$ and $E_{y,n} = (n + 1/2)\hbar\omega$, so the total energy is $E_{m,n} = (m + n + 1)\hbar\omega$. All the energy levels apart from the ground state are degenerate. For example the two different states

$$f^0(x, y) = Ax e^{-x^2/2a^2} e^{-y^2/2a^2}, \qquad g^0(x, y) = Ay e^{-x^2/2a^2} e^{-y^2/2a^2} \tag{81}$$

where $a = (\hbar/m\omega)^{1/2}$ and $A = \sqrt{2/\pi a^4}$ both have energy $E_{1,0} = E_{0,1} = 2\hbar\omega$.

Suppose now a perturbing potential of the form

$$V' = \epsilon(x + y)^2$$

is applied to the system (for example, if the system is an electron in a trap, this perturbation could be caused by an electric field along the diagonal $x = y$ and increasing linearly with distance). It is quite easy to see what must happen: the potential energy is no longer circularly symmetric, the well now has an elliptical shape, being more tightly confining along the diagonal $x = y$.

In order to analyze the problem, it is convenient to introduce new coordinates $x' = (x+y)/\sqrt{2}$, $y' = (x - y)/\sqrt{2}$, so the total potential energy can be written

$$V_0 + V' = \frac{1}{2}m\omega^2(x^2 + y^2) + \epsilon(x + y)^2 = \left(\frac{1}{2}m\omega^2 + 2\epsilon\right) x'^2 + \frac{1}{2}m\omega^2 y'2$$

Define ω' by putting $(1/2)m\omega^2 + 2\epsilon = (1/2)m\omega'^2$ and we obtain the exact solution by separation into two one-dimensional problems as before, only now in the new coordinate system x', y'. The energy levels are now $E = (m + 1/2)\hbar\omega' + (n + 1/2)\hbar\omega$ and the energy eigenstates are of the form $\psi = u(x + y)v(x - y)$.

The first significant point is that the new choice of coordinates was open to us all along: we could just as readily have written the eigenstates of the original unperturbed Hamiltonian in this coordinate system. We would have obtained for the first excited states, for example,

$$\psi_+^0(x,y) = A\frac{x+y}{\sqrt{2}}e^{-(x+y)^2/4a^2}e^{-(x-y)^2/4a^2} = \frac{1}{\sqrt{2}}(f^0(x,y) + g^0(x,y)) \tag{82}$$

$$\psi_-^0(x,y) = A\frac{x-y}{\sqrt{2}}e^{-(x+y)^2/4a^2}e^{-(x-y)^2/4a^2} = \frac{1}{\sqrt{2}}(f^0(x,y) - g^0(x,y)) \tag{83}$$

$$\tag{84}$$

where the reader should check the final equality by expanding the squared terms in the exponents. The concept generalizes as follows: whenever we have degenerate states f^0, g^0 we can always combine them together in superpositions to get an equally valid alternative set of states, e.g. $(f^0 \pm g^0)/\sqrt{2}$.

Now let us examine how perturbation theory deals with this example. If we picked the original way of writing the states, $f^0(x,y)$ and $g^0(x,y)$, then we would have an infinity in equation (80) and therefore the original approximation in going from (74) to (75) has broken down. However, if we pick the equally valid zero-order states $\psi_+^0(x,y)$ and $\psi_-^0(x,y)$ then in equation (79) the left hand side, which is here $\langle\psi_+^0|\,\delta H\,|\psi_-^0\rangle$, is equal to zero, which is consistent with the right hand size, so everything is mutually consistent and correct, and the first order perturbation theory results are valid. To prove $\langle\psi_+^0|\,\delta H\,|\psi_-^0\rangle = 0$ simply examine the integral:

$$\langle\psi_+^0|\,\delta H\,|\psi_-^0\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A^*x'e^{-x'^2/2a^2}e^{-y'^2/2a^2}\,\epsilon x'^2\,Ay'e^{-x'^2/2a^2}e^{-y'^2/2a^2}\,dx'\,dy' = 0. \tag{85}$$

To confirm the perturbation theory value for the energy shifts, evaluate

$$\delta E_+ = \langle\psi_+^0|\,\epsilon x'^2\,|\psi_+^0\rangle = \frac{3\epsilon}{m\omega} \tag{86}$$

$$\delta E_- = \langle\psi_-^0|\,\epsilon x'^2\,|\psi_-^0\rangle = \frac{\epsilon}{m\omega} \tag{87}$$

$$\tag{88}$$

and compare with the exact expression expanded to first order in the perturbation size ϵ:

$$E_+ = \frac{3}{2}\hbar\omega' + \frac{1}{2}\hbar\omega \simeq 2\hbar\omega + \frac{3\epsilon}{m\omega} \tag{89}$$

$$E_- = \frac{1}{2}\hbar\omega' + \frac{3}{2}\hbar\omega \simeq 2\hbar\omega + \frac{\epsilon}{m\omega}. \tag{90}$$

9.2 The general method

We can now see how the method works in general:

Degenerate perturbation theory proceeds exactly as non-degenerate perturbation theory, as long as we ensure that the zero-order states satisfy $\langle \psi_n^0 | \delta H | \psi_k^0 \rangle = 0$ whenever $E_n^0 = E_k^0$.

When $\langle \psi_n^0 | \delta H | \psi_k^0 \rangle = 0$ we say the perturbation δH does not "connect" the two states.

At this stage you ought to be concerned that it may not always happen that $\langle \psi_n^0 | \delta H | \psi_k^0 \rangle = 0$ when $E_n^0 = E_k^0$, but we will show that in fact the degenerate states can always be manipulated into a form which ensures this. The manipulation is referred to as "diagonalizing the perturbing Hamiltonian", and can be achieved by an automatic though somewhat involved mathematical procedure. However, we can almost always avoid the full complexities of that procedure, by using methods which carry more physical insight. The methods to be tried first are

- make use of some symmetry of the problem, as in the example discussed in the previous section

- identify *constants of the motion*

9.3   Constants of the motion

The phrase 'constant of the motion' can be used in either classical or quantum mechanics, it refers to some physical property which does not change in time, although the physical system in question may be changing in time in other ways. For example, for a ball thrown across a tennis court, the horizontal component of momentum is a constant of the motion (in the absence of air resistance) while the vertical component of momentum, and height, and the horizontal position, are not. For a thrown gyroscope (or a spinning tennis ball) the angular momentum as well as the horizontal velocity is a constant of the motion.

In the quantum context, a constant of the motion is a physical observable $\hat{Q}$ such as position, momentum, angular momentum (orbital or spin), etc. whose value does not change in time for the system under consideration. That is, if the observable were to be measured, the set of values which could be obtained, and their probability distribution, are independent of the time at which the measurement is made. You should know by now that if an observable $\hat{Q}$ does not evolve in time, then its operator must commute with the Hamiltonian. But this in turn implies that the eigenstates of $\hat{Q}$ can simultaneously be eigenstates of the Hamiltonian and hence that any state of definite energy can also have a definite value of $\hat{Q}$, so the eigenvalues of $\hat{Q}$ are *good quantum numbers*.

term	meaning	mathematical implication
constant of the motion	quantity which does not vary in time	the operator commutes with the Hamiltonian: $[\hat{Q}, \hat{H}] = 0$
good quantum number	quantum number associated with an operator which is a constant of the motion	eigenstate of $\hat{Q}$ is simultaneously an eigenstate of $\hat{H}$

These concepts are useful in perturbation theory when we can identify an observable which is a constant of the motion under the original Hamiltonian H^0 and also under the perturbed Hamiltonian $H^0 + \delta H$. Suppose $\hat{Q}$ is such an observable. Then we know two useful things:

- the zeroth order energy eigenstates can be (they don't *have to* be, but they *can* be) eigenstates

of $\hat{Q}$:

$$\hat{Q}\psi_k^0 = q_k\psi_k^0 \quad \text{for energy eigenstates} \quad \psi_k^0$$

- $\hat{Q}$ commutes with the perturbation:

$$[\hat{Q}, \delta H] = 0 \quad \Rightarrow \hat{Q}\,\delta H = \delta H\,\hat{Q}$$

From these we can prove the crucial condition $\langle\psi_n^0|\,\delta H\,|\psi_k^0\rangle = 0$ as follows:

$$
\begin{aligned}
\langle\psi_n^0|\,\delta H\,|\psi_k^0\rangle &= \frac{1}{q_k}\langle\psi_n^0|\,\delta H q_k\,|\psi_k^0\rangle \\
&= \frac{1}{q_k}\langle\psi_n^0|\,\delta H \hat{Q}\,|\psi_k^0\rangle \quad [\text{ since } \psi_k^0 \text{ is eigenstate of } \hat{Q}\\
&= \frac{1}{q_k}\langle\psi_n^0|\,\hat{Q}\delta H\,|\psi_k^0\rangle \quad [\text{ since } \hat{Q} \text{ commutes with } \delta H\\
&= \frac{1}{q_k}\langle\psi_n^0|\,q_n\delta H\,|\psi_k^0\rangle \quad [\text{ Hermitian } \hat{Q} \text{ and } \psi_n^0 \text{ eigenstate of } \hat{Q}\\
&= \frac{q_n}{q_k}\langle\psi_n^0|\,\delta H\,|\psi_k^0\rangle \\
\Rightarrow \quad \langle\psi_n^0|\,\delta H\,|\psi_k^0\rangle\left(1 - \frac{q_n}{q_k}\right) &= 0
\end{aligned}
$$

therefore

$$\text{either} \quad \langle\psi_n^0|\,\delta H\,|\psi_k^0\rangle = 0 \quad \text{or} \quad q_n = q_k$$

i.e. the perturbation does not connect states of different $\hat{Q}$.

The method to apply degenerate perturbation theory by means of constants of the motion is as follows.

1. Find an observable which is a constant of the motion under both the original and the perturbing Hamiltonian. Call this observable $\hat{Q}$.

2. We know it is possible to find states which are eigenstates both of the zero-order Hamiltonian and of $\hat{Q}$, so find them.

3. As long as we use these zero-order states, the standard results of perturbation theory will apply without problem (as proved above).

Sometimes the states identified in step (2) still have the problem of simultaneous degeneracy with respect to both energy and $\hat{Q}$, in which case either a further constant of the motion must be found or else the brute force 'diagonalisation' procedure (which I have not described) has to be resorted to. Assuming this does not happen (it will not for the cases we meet in this introduction to atomic physics), the only subtlety is that the states to be identified in step (2) might not be exactly the same as the states you were used to for the physical system under consideration. For example, in the 2D harmonic oscillator, we started with states based on x and y axes, and then later decided that states based on $x + y$ and $x - y$ axes were better. In atomic physics it is usually the angular momentum which is the issue. For example in hydrogen we typically first reckon the energy eigenstates to have well-defined $\hat{l}^2$, $\hat{l}_z$, $\hat{s}^2$ and $\hat{s}_z$, so the good quantum numbers in addition to n are l, m_l, $s = 1/2$, $m_s = \pm 1/2$. However, the spin-orbit interaction to be discussed in chapter

11 will lead us to think about another possibility: we can combine states of given l, s but different m_l, m_s to get states which are still energy eigenstates, but which now have well-defined values for the total angular momentum $\hat{j}^2$ where $\hat{\mathbf{j}} = \hat{\mathbf{l}} + \hat{\mathbf{s}}$. Now j and m_j are good quantum numbers (as well as l and s) but m_l and m_s are not.

Since the case of coupled angular momentum is so prevalent in atomic physics, I will finish this chapter by discussing it in detail, first for classical mechanics, and then with the main quantum ideas.

9.4   Example: charged gyroscopes

Suppose we have two classical gyroscopes in a space station, so that they are in free fall and we can ignore gravity. The gyroscopes are spinning fast, and their axes are not parallel. Each gyroscope will have a fixed orientation in space because the angular momentum about any axis in space does not change in time (conservation of angular momentum). Now suppose that each gyroscope has a small lump of electric charge placed somewhere on the flywheel—think of an insulating plastic patch charged by a few electrons for example. The gyroscopes will now influence each other by Coulomb repulsion, and the exact form of this influence as a function of position and time is very complicated because the gyroscopes are rotating. However, imagine that the charge really is very small, so the influence of the gyroscopes on each other is tiny compared with the size of each one's rotation about its own axis. What will happen? We can make three predictions, two of them certain and exact, the other approximate but correct to a high degree of accuracy.

What is certain is that

1. *Something* will happen—the motion will change in some way.

2. The total angular momentum of both gyroscopes together will not change. This is certain because internal forces (i.e. internal to the system consisting of the pair of gyroscopes) cannot produce a net torque on the whole system.

A very reasonable (and in fact quite accurate) approximation is that the speed of rotation of each gyroscope about its own axis will not change. I am not going to prove this here, [Problem: (hard) find a way to treat the small extra torques which arise in this scenario and see if you can prove the claim], but I claim that it is physically reasonable. What happens is that both gyroscopes begin to precess slowly, and they do so in such a way that the total angular momentum of the whole system stays constant. That means they precess *about the direction of the total angular momentum*. Think about it!—draw a vector diagram of the two individual angular momentum vectors adding up to make the total angular momentum, and figure out a way to have the two vectors move with constant length but changing direction in such a way that their sum does not change.

In mathematical terms, the size of the two classical angular momenta l_1, l_2 does not change, while their changing direction means their individual z components fluctuate in time. Meanwhile the total angular momentum vector $\mathbf{L} = \mathbf{l}_1 + \mathbf{l}_2$ is fixed, i.e. fixed in both size and direction. $\mathbf{l}_1$ and $\mathbf{l}_2$ precess about $\mathbf{L}$.

9.5 Quantum gyroscopes

Now we will analyse the same problem using quantum mechanics. In order to have a reasonable picture, assume the gyroscopes are very small and rotating very fast, and the charge is just one electron's worth—in fact our two gyroscopes will be simply two electrons in an atom! This is a classic example of degenerate perturbation theory. Before we allow for the influence of the electron-electron repulsion, the energy eigenstates ψ_k^0 are angular momentum states having quantum numbers l_1, m_{l1}, l_2, m_{l2}. The fact that these are eigenstates of z-component of angular momentum $\hat{l}_{z1}$, $\hat{l}_{z2}$, as well as of total angular momentum squared for each electron, indicates that the gyroscopes have a fixed orientation in space. Next we allow for the perturbation caused by their Coulomb repulsion. We find that the energy eigenstates ψ_k of the complete Hamiltonian $H^0 + \delta H$ are still, to good approximation, eigenstates of $\hat{l}_1^2$, $\hat{l}_2^2$, but they are *no longer* eigenstates of $\hat{l}_{z1}$, $\hat{l}_{z2}$. Therefore the numbers m_{l1}, m_{l2} are not good quantum numbers any more. This expresses the fact that the z component of angular momentum of each electron now wanders up and down due to precession (see diagram in lecture). The energy eigenstates ψ_k now combine states of the same l_i but different m_{li}, so as to produce states of definite L and M_L. This corresponds to the statement that the total angular momentum $\hat{\mathbf{L}}$ remains constant under the perturbation, having both a constant size and a constant direction.

Actually, and this is important, we could have used states of definite L and M_L all along. That is, instead of picking ψ_k^0 to be a state given by quantum numbers l_1, m_{l1}, l_2, m_{l2}, we could have chosen it to be a state of definite l_1, l_2, L, M_L, because that would have been an eigenstate of H^0 as well. This is the trick of degenerate perturbation theory: we have some choice about exactly which form the zero order states have, because we can take groups of degenerate states and combine together the members within in each group. If we can do this in such a way as to obtain degenerate states which are **not connected to one another by the perturbation**, then our work is done since the blowing-up problem of equation (80) is avoided, and we simply use the standard results (77), (80) for the energy shift and change in wavefunction.

In this example the perturbation can't connect states of different L, M_L, so these are the states which the system naturally adopts as its energy eigenstates (recall the behaviour of the 2D harmonic oscillator and of the vibrating drum skin) and each acquires a well-defined energy shift. The energy shift is given by

$$\delta E = \left\langle \{q.\}, L, M_L \left| \delta H \right| \{q.\}, L, M_L \right\rangle \tag{91}$$

where $\{q.\}$ is the list of all the other relevant quantum numbers, which in this case is n_1, n_2, l_1, l_2 but N.B. not m_{l1}, m_{l2}.

10 Russell-Saunders or LS coupling

We are now ready to finish accounting for electrostatic interactions in atoms. So far we have treated the central field approximation, in which all the electrons are reckoned to move in a common spherically-symmetric field. The rest of their repulsion, the part which is not spherically symmetric, is called the *residual electrostatic interaction.*

It is easy to write down the mathematical form of this residual electrostatic interaction. The precise complete Hamiltonian for the atom (ignoring small spin-related effects which we haven't considered yet) is

$$H = \sum_i \left(\frac{\mathbf{p}_i^2}{2m} - \frac{Ze^2}{4\pi\epsilon_0 r_i} \right) + \sum_{i>j} \frac{e^2}{4\pi\epsilon_0 r_{ij}} \tag{92}$$

while the approximate Hamiltonian used in the central field approximation was

$$H_{\text{central}} = \sum_i \left(\frac{\mathbf{p}_i^2}{2m} + V(r_i) \right) \tag{93}$$

(recall section 7). Therefore the residual electrostatic interaction is

$$H_{\text{re}} = H - H_{\text{central}} = \sum_i \left(-\frac{Ze^2}{4\pi\epsilon_0 r_i} - V(r_i) \right) + \sum_{i>j} \frac{e^2}{4\pi\epsilon_0 r_{ij}} \tag{94}$$

Needless to say, this is a highly complicated Hamiltonian! However, we can work out the main features of its effects using a very simple physical insight: whatever else H_{re} does, it cannot change the total orbital angular momentum of all the electrons. Therefore $[\hat{\mathbf{L}}, H_{\text{re}}] = 0$ where

$$\hat{\mathbf{L}} = \sum_i \hat{\mathbf{l}}_i \tag{95}$$

and we know immediately that the atomic energy eigenstates, under the influence of H_{re}, are eigenstates of $\hat{L}^2$ and $\hat{L}_z$! Also, states of different L will now have different energy. The effect therefore is to *split* the energy associated with any given configuration into a set of *terms* of given L.

We should now recall the discussion of the influence of exchange symmetry and spin. Only exchange-antisymmetric total states are allowed, and when we combine the spatial and the spin parts, this means that for a given L state (spatial), the symmetry results in spin states of definite total spin S, with states of different S (spin) being associated with different spatial wavefunctions, and hence having different energies even though the interaction in question has nothing to do with spin.

The upshot is that the quantum numbers on which the energy of the whole atom depends are now n_i, l_i, L, S, and this whole method of approximation is called *LS coupling* or Russell-Saunders coupling. The things which are coupled in this coupling are:

1. All the individual orbital angular momenta couple together to make $\hat{\mathbf{L}}$.

2. All the individual spin angular momenta couple together to make $\hat{\mathbf{S}}$.

We can picture all the individual orbital angular momenta $\mathbf{l}_i$ precessing about $\mathbf{L}$, and all the all the individual spin angular momenta $\mathbf{s}_i$ precessing about $\mathbf{S}$.

LS coupling is most easily understood by looking at a few examples. We have to recall the rules for adding angular momenta (section 5).

config.	possible terms	notation	num. of terms	num. of states
1s	$L = 0$, $S = 1/2$	1s ^2S	1	2
2p	$L = 1$, $S = 1/2$	2p ^2P	1	6
1s2s	$L = 0$, $S = 0$ or 1	1s2s ^1S and ^3S	2	4
1s2p	$L = 1$, $S = 0$ or 1	1s2p ^1P and ^3P	2	12
2p3p	$L = 0$ or 1 or 2, $S = 0$ or 1	2p3p ^1S, ^3S, ^1P, ^3P, ^1D, ^3D	6	36

In this table we have introduced the term notation

$$^{2S+1}L$$

where the superscript indicates the value of S by giving the multiplicity (number of M_S states for the given S) and the capital letter gives the value of L by the usual convention. Note that there is now an annoying ambiguity about whether the letter S refers to $L = 0$ or to the spin, but we just have to live with that and make sure we are clear in practice.

The final column in the table is instructive. It gives the total number of quantum states in the configuration, which we met before when considering the periodic table. Thus for 1s2p for example there are two spin states for each electron, and one orbital angular momentum state for the first electron, 3 for the second, making a total of $2 \times 2 \times 1 \times 3 = 12$ states. It is very nice to notice that the 2 terms of this configuration, ^1P and ^3P, together exactly account for all the 12 states: the ^1P term has 1 spin state and 3 orbital states; the ^3P term has three of each, therefore the total number of states is $1 \times 3 + 3 \times 3 = 12$, in agreement with the first calculation. [Problem: confirm the final entry in the table (36 states for the configuration $2p3p$), both for the individual electron angular momentum states, and for the terms.]

N.B. there is a special case we left out of the table, namely the case when both electrons have the same spatial state, such as $1s^2$ or $2p^2$. In this case some of the terms are missing because the total angular momentum states of the atom can now only be antisymmetric. For the case ns^2 only the ^1S term is allowed; for np^2 only the terms ^1S, ^3P, and ^1D are allowed.

You should be aware that when the LS coupling approximation is discussed, (such as in finals, for example), a discussion of the fine structure is usually also included, which we haven't covered yet. Therefore you should regard the LS coupling approximation as not quite finished with yet—we need to add the fine structure, which splits each *term* into one or more *levels*. Also, when we consider this, we will find that in fact the LS coupling approximation **does not work for all atoms**. It works well for light atoms, especially the ones in groups 1 and 2, which are the ones we concentrate on when learning atomic physics. We shall see why this is, and why it doesn't work for some other atoms, in section 12.

Finally, did you notice that we did some degenerate perturbation theory just now? —we identified the constants of the motion $\hat{\mathbf{L}}$ and $\hat{\mathbf{S}}$ and therefore talked about atomic states of definite L, S, M_L, M_S. We then chose not to trouble to work out the size of the energy shifts, except to note that we expect them to be small compared to the gross structure—though not very small, since it is still the electrostatic interaction which is being considered. Actually, we can say something about the shifts: we expect symmetric spatial wavefunctions to have higher energy than their antisymmetric partners, and therefore singlets to have more energy than the corresponding triplet for atoms in group 2 (same argument already given in detail for helium). This rule works quite well when only one electron has $l > 0$, but with more than 2 electrons in the atom the relative energies of the totally symmetric and antisymmetric spatial states can come out either way, so triplets are sometimes below, sometimes above the corresponding singlets.

10.1 Main points

The main points of this chapter have been:
the residual electrostatic interaction; fact that total $\mathbf{L}$ and total $\mathbf{S}$ are constants of the motion; vector model of the coupling; term notation, and the set of terms which arise from a given configuration; order of magnitude of the energy separation of the terms.

11 Fine structure

It should not come as a surprise that our treatment of hydrogen so far has 'only' been accurate to a level of $\alpha^2 \simeq 5 \times 10^{-5}$. This is because the electron is moving at speeds of order αc, so relativistic effects of order α^2 are going to be present. In many branches of physics a precision better than one in a thousand would already be regarded as very good, but here at the fundamentals we have to do better than that. Furthermore, our treatment of other atoms such as sodium has been somewhat less accurate still, partly because the electron moves quicker when it is near a more highly-charged nucleus. As for hydrogen-like Uranium, $_{92}U^{91+}$, our treatment is essentially unusable because the electron is moving at speeds of order half the speed of light. [Problem (easy): use the hydrogenic scaling law to confirm this statement.]

11.1 The Dirac result

To cope with relativity (to be technically precise, to cope with Lorentz covariance) properly we have to reject the Schrödinger equation altogether, and replace it with an equation which behaves the right way under changes of reference frame. Such an equation was discovered by Dirac. It had all the right properties, only it had the surprising feature that instead of treating one wavefunction at a time, it used a vector of four wavefunctions to describe a single particle! The genius of Dirac was not only to discover the equation, but also to reason that the four-wavefunction vector described the two spin states of the electron, and predicted a previously unknown form of the electron, which we now know as its anti-particle, the positron.

The Dirac equation for two point-particles interacting by a $1/r$ potential can be solved exactly, if we assume one particle is fixed (i.e. it has infinite mass), and the answer for hydrogen is

$$E_{nj} = mc^2 \left[1 + \alpha^2 \left(\frac{1}{n - (j + \frac{1}{2}) + [(j + \frac{1}{2})^2 - \alpha^2]^{1/2}} \right)^2 \right]^{-1/2} \tag{96}$$

where j is the total angular momentum quantum number ($\hat{\mathbf{j}} = \hat{\mathbf{l}} + \hat{\mathbf{s}}$) for the electron. To allow for a nucleus of charge Ze, just replace α by $Z\alpha$. Note that this energy includes the rest energy mc^2, so if we want to recover the Schrödinger prediction we should expand the square root in powers of α, and then take off the mc^2 contribution. [Problem: confirm that you then get the hydrogen energy $(-(1/2)m\alpha^2 c^2/n^2)$ to 2nd order in α].

The Dirac equation itself is beyond this level of the course, but you should be aware that it is the right relativistic equation, and you should notice two things about the prediction for the energy levels:

1. The formula contains α only in the form α^2, so when expanded in a power series, the next term after the hydrogen gross structure $(-(1/2)m\alpha^2 c^2)$ will be of order α^4. Allowing for a nucleus of charge Z, this becomes order $Z^4 \alpha^4$.

2. The formula depends on j, and therefore predicts slightly different energies for two states of the same n but different j. This difference is called *fine structure*.

I can't resist adding that even the Dirac equation is slightly off: in 1947 Lamb and Retherford[10] (not

---

[10]W. E. Lamb and R. C. Retherford, Physical Review **72**, 241(1947).

Rutherford), in a triumph of experimental physics, precisely measured a small splitting between states of different l but the same j, which is not predicted by Dirac on its own. The splitting is now called the Lamb shift (another Nobel prize here). This greatly stimulated the discovery of Quantum Electrodynamics which accounts for this further effect and pretty much everything else about electrons and photons.

11.1.1 Dirac summary

The summary of the Dirac equation prediction, up to order $Z^4\alpha^4$, is

power of $Z\alpha$	coefficient	interpretation
	mc^2	rest energy
$Z^2\alpha^2$	$-mc^2/2n^2$	gross structure
$Z^4\alpha^4$	$+\frac{mc^2}{8n^4}(3-\frac{4n}{j+1/2})$	fine structure

This power series gives an easy way to remember that the fine structure in hydrogen-like ions scales as Z^4.

11.2 Hydrogen fine structure according to Schrödinger

Although we have said that the Schrödinger equation is not quite perfect when it comes to understanding relativistic effects, nevertheless we can use it by treating the relativistic effects as small perturbations (ok for hydrogen, but not ok for hydrogen-like uranium). This gives a lot of physical insight which we can then apply to other atoms.

There are 3 corrections:

1. Relativistic kinetic energy correction ('mass correction').

$$\begin{aligned} T &= (p^2c^2 + m^2c^4)^{1/2} - mc^2 \quad [\text{ relativistic energy} \\ &\simeq \frac{p^2}{2m} - \frac{1}{2mc^2}\left(\frac{p^2}{2m}\right)^2 + \ldots \end{aligned}$$

And we can sneakily express the kinetic energy $p^2/2m$ as total energy minus potential energy, so the correction Hamiltonian is

$$H_m = -\frac{1}{2mc^2}(E_n - V)^2. \tag{97}$$

2. Spin–orbit interaction
This is the interaction between the magnetic dipole moment of the electron and the magnetic field it experiences owing to its motion through the electric field of the nucleus. (In other terms, one may say that the nucleus moves with respect to the electron and so produces a magnetic field at the position of the electron.) We will calculate the contribution by an argument which is fairly simple and has good physical insight but which unfortunately needs fixing at the end by a factor of two. The appendix gives a calculation which is more complicated, but which has the merit of being exact and so proves the fix-up needed in the first method.

To understand magnetic dipoles, first let's examine how they come about in classical physics. For example, consider a classical electron moving in a circular orbit of radius r. If it has speed v then

the orbital angular momentum $l = rmv$ and the current around the loop is $ev/2\pi r$. The magnetic dipole moment of a current loop is the current times the area, so

$$|\mu| = \frac{\pi r^2 ev}{2\pi r} = \frac{e}{2} rv = \frac{e}{2m} l \tag{98}$$

Examining the signs and directions carefully, we find that the full relationship is

$$\boldsymbol{\mu} = -\frac{e}{2m} \mathbf{l}, \tag{99}$$

so we immediately have the quantum operator relationship

$$\hat{\boldsymbol{\mu}} = -\frac{e}{2m} \hat{\mathbf{l}} \tag{100}$$

for the magnetic moment associated with the orbital angular momentum of an electron.

It is useful to gather all the dimensional quantities on the right hand side together in one place, giving us a unit of magnetic dipole moment called the *Bohr magneton*:

$$\mu_B = \frac{e\hbar}{2m}. \tag{101}$$

We can expect spin angular momentum also to have a magnetic moment associated with it, and it does, but it turns out that the proportionality constant is twice that for orbital angular momentum (Dirac's equation predicts this[11], I know of no simpler argument for it). To allow for all eventualities we therefore introduce a dimensionless factor called the g factor, and write:

$$\boldsymbol{\mu} = -g_j \mu_B \frac{\hat{\mathbf{j}}}{\hbar}, \tag{102}$$

with

$$
\begin{aligned}
g_l &= 1 \\
g_s &= 2.
\end{aligned}
$$

We will now introduce a new convention. Up till now we have used the symbols $\hat{\mathbf{j}}$, $\hat{\mathbf{l}}$ and $\hat{\mathbf{s}}$ for the angular momenta, so that they are dimensional quantites. However, it will become more and more convenient to take the angular momentum units into account by defining new symbols

$$
\begin{aligned}
\hat{\mathbf{j}}' &= \hat{\mathbf{j}}/\hbar \tag{103} \\
\hat{\mathbf{l}}' &= \hat{\mathbf{l}}/\hbar \tag{104} \\
\hat{\mathbf{s}}' &= \hat{\mathbf{s}}/\hbar \tag{105}
\end{aligned}
$$

so that, for example, equation (102) becomes

$$\boldsymbol{\mu} = -g_j \mu_B \hat{\mathbf{j}}' \tag{106}$$

and

$$(\hat{\mathbf{l}}')^2 Y_{l,m} = l(l+1) Y_{l,m}. \tag{107}$$

---

[11]In fact Quantum Electrodynamics makes a slight adjustment, and the true g factor for spin is greater than 2 by an amount of order α, see formula in section 1.2.

Next, and this is the confusing bit, we will begin to adopt the very common practice of dropping the prime on $\hat{\mathbf{j}}'$, $\hat{\mathbf{l}}'$, $\hat{\mathbf{s}}'$ $\hat{\mathbf{L}}'$, etc. You can always tell which convention is being used by checking the dimensions, but **make sure you know which convention has been adopted, and that your own work is consistent**.

The above serves as an introduction to some notations, and a reminder on magnetic dipoles. Now we will consider the magnetic field experienced by the electron in hydrogen. One way to think about this is to say that the electron sees the nuclear charge going around it, so the nuclear charge q moving at velocity $\mathbf{v}'$ with respect to the electron produces a magnetic field at the electron according to the Biot-Savart law:

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi}\frac{q\mathbf{v}' \wedge \mathbf{r}}{r^3} \tag{108}$$

$$= \mu_0\epsilon_0\frac{q}{4\pi\epsilon_0}\frac{\mathbf{v}' \wedge \hat{\mathbf{r}}}{r^2} \tag{109}$$

$$= \frac{-\mathbf{v} \wedge \mathbf{E}}{c^2} \tag{110}$$

where $\mathbf{E}$ is the electric field of the nucleus at the position of the electron, and $\mathbf{v}$ is the velocity of the electron in the reference frame of the nucleus. Here is the relativistically insightful way to think about equation (110). You are used to the idea that when a particle moves in a magnetic field it experiences a force proportional to its charge and to $\mathbf{v} \wedge \mathbf{B}$. Such a force is equal to that which would be produced by an electric field of size and direction $\mathbf{v} \wedge \mathbf{B}$. Equation (110) is the corresponding result for a particle moving through an electric field: it experiences forces equal to those which would be produced by the magnetic field $-\mathbf{v} \wedge \mathbf{E}/c^2$. The Biot-Savart formula is an approximation to the full relativistic result, being valid up to first order in v/c.

For an atom of many electrons, any given electron experiences a magnetic field from the nucleus and all the other electrons which move with respect to it. Each of the contributions has the form (110) with $\mathbf{E}$ being the contribution to the electric field at the electron in question, so the total magnetic field is still given by (110), with $\mathbf{E}$ now the total electric field at the electron in question. We can also see this immediately by regarding (110) as the relativistic formula (to first order in velocity) for the transformation of the fields when going from one reference frame to another.

Notice that for the case of a radial field, $\mathbf{E}$ is in the direction of $\mathbf{r}$ and therefore $\mathbf{B}$ is in the direction of $-\mathbf{v} \wedge \mathbf{r} = \mathbf{l}/m$, i.e. in the direction of the orbital angular momentum vector.

This magnetic field interacts with the magnetic dipole moment of the electron,

$$H_{so} = -\boldsymbol{\mu} \cdot \mathbf{B}$$

$$= g_s\frac{e\hbar}{2m}\hat{\mathbf{s}}' \cdot \frac{-\mathbf{v} \wedge \mathbf{E}}{c^2} \qquad [\text{ where } \mathbf{s}' \text{ is dimensionless}$$

$$= \frac{\hbar}{mc^2}\hat{\mathbf{s}}' \cdot \left(\frac{1}{r}\frac{dV}{dr}\mathbf{r} \wedge \mathbf{v}\right) \qquad [\text{ where } V \text{ is potential energy}$$

$$\stackrel{?}{=} \frac{\hbar^2}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\hat{\mathbf{l}}' \cdot \hat{\mathbf{s}}' \qquad [\text{ where } \mathbf{l}' \text{ and } \mathbf{s}' \text{ are dimensionless; } \mathbf{r} \wedge m\mathbf{v} = \hbar\mathbf{l}$$

This calculation is correct but the question-mark highlights a subtlety: we have just done a calculation in the reference frame of the electron, which owing to the orbital motion is not an inertial reference frame. When we transform back to the centre of mass frame using special relativity, a further precession is introduced, called the *Thomas precession*, and the result in the centre of mass

frame is exactly half what we have calculated:

$$H_{so} = \frac{1}{2}\frac{\hbar^2}{m^2c^2}\frac{1}{r}\frac{dV}{dr}\hat{\mathbf{l}}\cdot\hat{\mathbf{s}} \quad [\text{ correct result} \tag{111}$$

$$= \frac{1}{2}\frac{\hbar^2 e^2}{4\pi\epsilon_0 m^2 c^2}\frac{1}{r^3}\hat{\mathbf{l}}\cdot\hat{\mathbf{s}} \quad [\text{ for } V = -e^2/4\pi\epsilon_0 r. \tag{112}$$

where now we are adopting the practice of dropping primes: the angular momenta are understood to be in units of $\hbar$. The first expression, (111) is more general because it is true for any central potential. The second form shows what happens in the case of a $1/r$ potential, it is worth noting that the interaction is then proportional to $1/r^3$.

Now let's see what happens when we apply this interaction to the hydrogen atom using perturbation theory. We have an interaction between $\hat{\mathbf{l}}$ and $\hat{\mathbf{s}}$ but we can be sure this will not manage to change the total angular momentum $\hat{\mathbf{j}} = \hat{\mathbf{l}} + \hat{\mathbf{s}}$ because, once again, the net angular momentum can't be affected by internal forces. Therefore the quantities $\hat{\mathbf{j}}^2$ and $\hat{j}_z$ are constants of the motion. Furthermore, we can express $\hat{\mathbf{l}}\cdot\hat{\mathbf{s}}$ in terms of $\hat{\mathbf{j}}^2$, $\hat{\mathbf{l}}^2$ and $\hat{\mathbf{s}}^2$ quite easily:

$$\hat{\mathbf{j}}^2 = \hat{\mathbf{j}}\cdot\hat{\mathbf{j}} \tag{113}$$

$$= \left(\hat{\mathbf{l}}+\hat{\mathbf{s}}\right)\cdot\left(\hat{\mathbf{l}}+\hat{\mathbf{s}}\right) \tag{114}$$

$$= \hat{\mathbf{l}}^2 + \hat{\mathbf{s}}^2 + \hat{\mathbf{l}}\cdot\hat{\mathbf{s}} + \hat{\mathbf{s}}\cdot\hat{\mathbf{l}} \tag{115}$$

$$= \hat{\mathbf{l}}^2 + \hat{\mathbf{s}}^2 + 2\hat{\mathbf{l}}\cdot\hat{\mathbf{s}} \quad [\text{ since } \hat{\mathbf{s}} \text{ commutes with } \hat{\mathbf{l}}$$

$$\implies \hat{\mathbf{l}}\cdot\hat{\mathbf{s}} = \frac{1}{2}\left(\hat{\mathbf{j}}^2 - \hat{\mathbf{l}}^2 - \hat{\mathbf{s}}^2\right) \tag{116}$$

This is an exact relationship between the operators[12].

It is now easy to apply the perturbation theory: we just take the expectation value of $\hat{\mathbf{l}}\cdot\hat{\mathbf{s}}$ in a state of well-defined j, l and s which using (116) obviously gives

$$\langle l,s,j|\hat{\mathbf{l}}\cdot\hat{\mathbf{s}}|l,s,j\rangle = \frac{1}{2}\left(j(j+1) - l(l+1) - s(s+1)\right). \tag{117}$$

The vector model picture of this coupling is that $\mathbf{l}$ and $\mathbf{s}$ are precessing about $\mathbf{j}$.

Applying this to hydrogen, i.e. combining (112) and (117), we obtain

$$\delta E_{so} = \langle n,l,s,j|H_{so}|n,l,s,j\rangle = \frac{1}{2}\frac{\hbar^2 e^2}{4\pi\epsilon_0 m^2 c^2}\left\langle\frac{1}{r^3}\right\rangle\frac{1}{2}\left(j(j+1) - l(l+1) - s(s+1)\right),$$
$$\tag{118}$$

and more generally,

$$\delta E_{so} = \langle n,l,s,j|H_{so}|n,l,s,j\rangle = \frac{1}{2}\frac{\hbar^2}{m^2 c^2}\left\langle\frac{1}{r}\frac{dV}{dr}\right\rangle\frac{1}{2}\left(j(j+1) - l(l+1) - s(s+1)\right) \tag{119}$$

for an electron moving in any spherically symmetric potential.

Since the energy of the levels now depends on j, we extend the term notation to indicate j as a subscript, thus

$\{L, S, J\}$ are indicated by $^{2S+1}L_J$

---

[12]Some textbooks quote a cosine rule involving $|\mathbf{l}||\mathbf{s}|\cos\theta$, but this is an unnecessary departure from strictly accurate quantum theory.

For example, the ground state of hydrogen is 1s $^2S_{1/2}$; the three lowest states of helium are 1s^2 $^1S_0$, 1s2s $^1S_0$ and 1s2s $^3S_1$. (For hydrogen and the alkalis the total L, S, J equal the individual l, s, j of the single valence electron, but for atoms with more than one valence electron this need not be the case.)

Finally, let us evaluate $\langle 1/r^3 \rangle$ in hydrogen. Recalling $a_0 m\alpha c = \hbar$ for the Bohr radius a_0, we can see that $\langle 1/r^3 \rangle$ should be proportional to $1/a_0^3 = m^3\alpha^3 c^3/\hbar^3$ and hence

$$\delta E_{so} = \langle H_{so} \rangle \propto \frac{1}{2}\frac{\hbar^2 e^2}{4\pi\epsilon_0 m^2 c^2}\frac{m^3\alpha^3 c^3}{\hbar^3} = \frac{1}{2}mc^2\alpha^4. \tag{120}$$

The full result is

$$\left\langle \frac{1}{r^3} \right\rangle = \frac{1}{a_0^3 l(l+1/2)(l+1)n^3} \qquad [\text{ for } l > 0 \tag{121}$$

so

$$\delta E_{so} = \frac{mc^2\alpha^4}{2n^3}\frac{1}{l(l+1/2)(l+1)}\left\langle \hat{\mathbf{l}}\cdot\hat{\mathbf{s}} \right\rangle \tag{122}$$

for $l > 0$. Note that the dependence on $1/r^3$ leads to a dependence on $1/n^3$. When the spin-orbit interaction is combined with the effect of the mass correction, we obtain the correct Dirac result to order α^4, except that we have a difficulty when $l = 0$ since for that case we have a 'zero divided by zero' result.

3. Darwin term for $l = 0$

The final relativistic correction to the Schrödinger equation for hydrogen is a special case which only applies to $l = 0$. It is to do with the fact that the wavefunction for $l = 0$ does not vanish at the origin, so the non-relativistic approximation $Ze^2/4\pi\epsilon_0 r \ll mc^2$ is seriously violated. The interaction Hamiltonian is

$$H_D = -\frac{\hbar^2}{4m^2 c^2}e\mathbf{E}\cdot\boldsymbol{\nabla} \tag{123}$$

and the associated energy shift is proportional to $|\psi(0)|^2$ (the electron probability density at the nucleus). By a strange luck the energy shift due to the Darwin term for $l = 0$ is exactly given by substituting $l = 0$ in the formula for the energy shift given by the sum of the two previous interactions! (kinetic energy correction and spin–orbit interaction.) Actually this luck is not so strange since this whole perturbative calculation is somewhat artificial, introducing relativistic phenomena such as spin and mass energy without bothering to make the fundamental equation (Schrödinger) relativistic. We are bound to keep introducing terms until we get the correct Dirac result!

Now, the above details of the Darwin term are off the syllabus, but as a matter of general knowledge you should know that
(1) the states of the same j (and n) are degenerate in hydrogen according to all but QED
(2) the spin-orbit shift is zero for $l = 0$
(3) the mass shift plus spin-orbit is the whole story for $l \neq 0$, but not for $l = 0$.

Adding the contributions of the three terms outlined above, we get the Dirac energy correct to order $Z^4\alpha^4$, see the Dirac summary 11.1.1.

You should check your understanding by working out the size of the fine structure from the formula. The observed fine structure of hydrogen is given in many textbooks. To have an order of magnitude in mind, notice that the splitting of the $2p\,^2P_{3/2}$ level from the $2p\,^2P_{1/2}$ level is about 10 GHz.

11.3 Main points

The main points of this chapter have been:
the fact that the Dirac equation gives a full relativistic treatment; the formula for the energy levels can be written as a power series in $Z^2\alpha^2$; hence the Z^4 scaling of fine structure; the Schrödinger treatment by perturbation theory with three contributions: kinetic energy correction, spin-orbit interaction, Darwin term for $l = 0$; the relationship between spin and magnetic moment; gyromagnetic ratio g and Bohr magneton μ_B; the magnetic field experienced by an electron in an atom; the general form of δE_{so}.

11.4 Appendix: off-syllabus exact calculation of hydrogen spin-orbit coupling

We can prove (111) either by working out the special relativity of a spinning object moving fast around an orbit, as Thomas did (look it up if you are interested), or by doing the calculation from the start in the rest frame of the nucleus, which is (to very good approximation) the centre of mass frame. In the latter case, we argue that the magnetic dipole moment of the electron (from its spin) has an associated magnetic field, and we consider this field at the nucleus. The field varies in time, because of the motion of the electron, and therefore by Maxwell's equation $\boldsymbol{\nabla} \wedge \mathbf{E} = -\partial\mathbf{B}/\partial\mathbf{t}$ it produces an electric field at the nucleus, which interacts with the nuclear charge. This reversal of which particle produces which type of field, compared to the calculation in the electron's rest frame, is a nice example of relativity in action in electromagnetism.

The calculation is best done in terms of the vector potential, and then using $\mathbf{E} = -\partial\mathbf{A}/\partial t$.

If the electron has magnetic moment $\boldsymbol{\mu}$ and position vector $\mathbf{r}$, then the vector potential *at the nucleus, due to the electron* is

$$\mathbf{A} = -\frac{\mu_0}{4\pi}\frac{\boldsymbol{\mu}\wedge\mathbf{r}}{r^3} \qquad [\,\mathbf{r} \text{ is position of electron}$$

$$\implies \mathbf{E} = -\frac{\partial\mathbf{A}}{\partial t} = \frac{\mu_0}{4\pi}\left(\frac{\dot{\boldsymbol{\mu}}\wedge\mathbf{r}}{r^3} + \frac{\boldsymbol{\mu}\wedge\mathbf{v}}{r^3} - 3\frac{\boldsymbol{\mu}\wedge\mathbf{r}}{r^4}\dot{r}\right) \qquad [\,\text{field at nucleus due to moving spin}$$

$$\implies H_{so} = -\int_\infty^r \mathbf{F}\cdot d\mathbf{r} \qquad [\,\text{force equals} -\text{gradient of energy}$$

$$= -Ze\frac{\mu_0}{4\pi}\int_\infty^r \frac{\boldsymbol{\mu}\wedge\mathbf{v}}{r^3}\cdot d\mathbf{r}$$

$$= Ze\frac{\mu_0}{4\pi}\frac{\boldsymbol{\mu}\wedge\mathbf{v}\cdot\mathbf{r}}{2r^3}$$

$$= -\frac{Ze}{4\pi\epsilon_0 c^2}\frac{\boldsymbol{\mu}\cdot\mathbf{r}\wedge\mathbf{v}}{2r^3} \qquad [\,\boldsymbol{\mu} = -g\mu_b\mathbf{s} = -2(e\hbar/2m)\mathbf{s}$$

$$= \frac{Ze^2}{4\pi\epsilon_0 r^3 c^2}\frac{\mathbf{s}\cdot\mathbf{l}\hbar^2}{2m^2} \qquad [\,\hbar\mathbf{l} = \mathbf{r}\wedge\mathbf{v}m$$

$$= \frac{\hbar^2}{2m^2c^2}\left(\frac{1}{r}\frac{dV}{dr}\mathbf{s}\cdot\mathbf{l}\right) \qquad [\,\text{this proves eq. (111)}$$

12 Fine structure for many electrons

We will now study fine structure in a many-electron atom. The central field approximation is the correct starting point for all atoms, and after that we have both the residual electrostatic interaction and the spin-orbit interaction:

$$H = H_{cf} + H_{re} + H_{so} \tag{124}$$

where we already met the central field and residual electrostatic Hamiltonians, and for many electrons

$$H_{so} = \sum_i \eta_i \hat{\mathbf{l}}_i \cdot \hat{\mathbf{s}}_i. \tag{125}$$

where we have taken all the constants and r-dependence into η_i.

There are three cases to consider:
1. $\langle H_{re} \rangle \gg \langle H_{so} \rangle$ $\rightarrow LS$ coupling.
2. $\langle H_{re} \rangle \ll \langle H_{so} \rangle$ $\rightarrow jj$ coupling.
3. $\langle H_{re} \rangle \simeq \langle H_{so} \rangle$ $\rightarrow$ no simple coupling scheme.
In case (1) we take H_{re} as the first perturbation to the central field, and then H_{so} after that, while in case (2) we take H_{so} as the first perturbation, and then H_{re} afterwards. In case (3) both Hamiltonians have to be taken together as a single perturbation to H_{cf}.

In each case we are doing degenerate perturbation theory, so we need simply identify which zeroth-order states $\left| \psi^0 \right\rangle$ are eigenstates of the perturbation, and the first order energy shift will be $\left\langle \psi^0 \right| H' \left| \psi^0 \right\rangle$.

12.1 1. $\langle H_{re} \rangle \gg \langle H_{so} \rangle$: LS coupling

This is typical for light atoms. The residual electrostatic interaction causes all the $\mathbf{l}_i$ to precess about $\mathbf{L}$, and combined with wavefunction symmetry this means all the $\mathbf{s}_i$ precess about $\mathbf{S}$. This means that all the individual $\hat{\mathbf{j}}_i^2$ of the electrons are not constants of the motion (in contrast to hydrogen), and the j_i are not good quantum numbers. Therefore the individual $\mathbf{l}_i \cdot \mathbf{s}_i$ are no longer constants of the motion. When we evaluate the perturbation caused by H_{so} we find that it acts instead as a coupling between the total $\mathbf{L}$ and $\mathbf{S}$ since they are the constants of the motion before the perturbation is applied. Mathematically, this is

$$\langle q; LSJ | \sum_i \eta_i \hat{\mathbf{l}}_i \cdot \hat{\mathbf{s}}_i | q; LSJ \rangle = \langle q; LSJ | A(q) \hat{\mathbf{L}} \cdot \hat{\mathbf{S}} | q; LSJ \rangle \tag{126}$$

where A is a proportionality constant which depends on all the other quantum numbers q, which are here $\{n_i, l_i\}$ (all the s_i are always 1/2 of course), i.e. the ones given by the configuration.

In the vector model, what is happening is that all the $\mathbf{l}_i$ precess rapidly about $\mathbf{L}$ and the $\mathbf{s}_i$ about $\mathbf{S}$, and meanwhile $\mathbf{L}$ and $\mathbf{S}$ precess slowly about $\mathbf{J}$ because of the spin-orbit coupling. We obtain

$$\delta E_{so} = \frac{1}{2} A \left(J(J+1) - L(L+1) - S(S+1) \right) \tag{127}$$

where the very simple maths is the same as seen previously in hydrogen.

In all LS coupled atoms apart from hydrogen and helium, this is the dominant contribution to the fine structure. In hydrogen the mass effect and Lamb shift also come in, and in helium some other

relativistic effects which take the mathematical form of orbit-orbit and spin-other-orbit ($\mathbf{s}_1 \cdot \mathbf{l}_2$) coupling are of the same order as spin-orbit.

The $J(J+1) - L(L+1) - S(S+1)$ form leads to the *Landé interval rule*, which states that the splitting between two fine structure levels of the same term (therefore the same configuration and L and S) is proportional to the larger J value:

$$\delta E_{so}(J) - \delta E_{so}(J-1) \propto J. \tag{128}$$

This is so easy to obtain from (127) that it is left as an exercise for the reader. [Answer: proportionality constant is A]. For an example, examine the splittings of the levels 3s3p $^3P_0$, $^3P_1$, $^3P_2$ in magnesium.

The constant A (i.e. constant for a given term, varying from one term to another) is related to the average $\langle (1/r) dV/dr \rangle$. For alkalis we get a dependence similar to hydrogen, with a modified Z and n. Since alkalis have spin 1/2 there are always two values of J for each L, namely $J = L \pm 1/2$, except for the S-states where $L = 0$ so $J = 1/2$ only. The energy levels are therefore split into doublets, and the splitting of the doublets is given approximately by

$$E_{\text{splitting}} = \alpha^2 \frac{Z_i^2 Z_o^2}{(n - \delta(l))^3 l(l+1)} E_R \tag{129}$$

where E_R is the Rydberg energy. You don't have to remember this formula, but it is interesting to note the $1/n^3$ dependence, and the fact that the hydrogenic $\alpha^4 Z^4$ is replaced by $\alpha^4 Z_i^2 Z_o^2$. Here, Z_i is an effective charge seen during the inner part of the orbit, which increases with Z, and Z_o is an effective charge seen during the outer part of the orbit, which equals 1 for neutral atoms.

In any atom (not just alkalis), $A > 0$ for less than half-filled subshells, so the lowest J lies lowest in energy, and the fine-structure *multiplet* is said to be *normal*. Otherwise it is said to be *inverted*. It is interesting to the following property also. The spin-orbit splitting does not move the 'centre of gravity' of the term. What we mean by this is that each J level has $2J+1$ states inside it, and the sum of the displacements of the levels multiplied by the $2J+1$ weights equals zero. For example, for $L = 1$, $S = 1$, the formula (127) says the $J = 2$ level is at $+A$, the $J = 1$ level at $-A$ and the $J = 0$ level at $-2A$. The 'centre of gravity' is therefore at $(5 \times 1 - 3 \times 1 - 1 \times 2)A = 0$ with respect to the unshifted term. [Problem: prove the result for any L, S.]

N.B. It is a very common mistake to confuse LS coupling with spin-orbit coupling! To avoid confusion, here they are compared and contrasted:

LS coupling, also called Russell-Saunders coupling, is a coupling scheme, that is, an approximate model of the structure of an atom	spin-orbit coupling is the interaction between the magnetic moment of the electrons and the magnetic field they experience owing to their motion through the electric field of the atom
in LS coupling the individual orbital angular momenta $\mathbf{l}_i$ couple together to make $\mathbf{L}$ and the individual spin angular momenta $\mathbf{s}_i$ couple together to make $\mathbf{S}$.	in spin-orbit coupling the spin angular momentum $\mathbf{S}$ couples to the orbital angular momentum $\mathbf{L}$
in an LS coupled atom, the spin-orbit interaction acts to couple $\mathbf{L}$ and $\mathbf{S}$, producing the fine structure	

Thus LS coupling is an angular momentum coupling scheme in which spin-orbit coupling plays a part, contributing the fine structure.

12.2 2. $\langle H_{so}\rangle \gg \langle H_{re}\rangle : jj$ **coupling**

This happens for some configurations in heavy atoms, since the high Z makes the spin-orbit coupling large. Recall that in hydrogen-like ions the spin-orbit coupling increases as Z^4, and the interaction is proportional to $\langle 1/r^3\rangle$ which emphasises small r, so in many-electron atoms the spin-orbit coupling is 'less screened' than the gross structure.

In jj coupling, then, the spin-orbit interaction dominates the residual electrostatic interaction. The spin-orbit interaction couples each electron's orbital and spin angular momenta, and each electron's total angular momentum $\mathbf{j} = \mathbf{l} + \mathbf{s}$ is a constant of the motion. Hence the good quantum numbers are j_i, on which the energy depends, and we can either take m_{ji} or (J, M_J) for the rest. Putting in the residual electrostatic interaction as a further perturbation causes the $\mathbf{j}_i$ to precess about $\mathbf{J}$, so m_{ji} are finally not good quantum numbers but M_J is.

The standard notation for the j_i of two electrons is

$$(j_1, j_2)$$

See the figure in the lecture for a comparison of LS and jj coupling.

12.3 Evidence for good LS coupling

The following signs, if they are present in any given atom, indicate that LS coupling is a good approximation:

1. The fine structure is small compared to the splitting between terms. This implies the heirarchy of perturbations is holding well.

2. The Landé interval rule is obeyed, except that this doesn't apply to the special case of helium which is LS coupled to good approximation although it doesn't obey the rule (owing to other spin-dependent interactions having similar strength to spin-orbit coupling in helium).

3. Strong transitions obey the selection rules $\Delta L = \pm 1, 0$ and $\Delta S = 0$.

4. The Zeeman effect agrees with the LS coupling formula for the Landé g_J factor.

The meaning of 3 and 4 will be clear later (see sections 14 and 15), I include them here simply to gather all the information into one place.

12.4 Main points

The main points of this chapter have been:
relative size of H_{re} and H_{so} determining the sequence of approximations, and hence the appropriate angular momentum coupling scheme; the vector model picture of LS coupling; the fine structure arising from spin-orbit coupling; Landé interval rule; fact that fine structure increases with atomic number (approximately as Z^2 for neutral atoms) and decreases with n and l; evidence for good LS coupling.

13 Hyperfine structure

We will now plunge another several orders of magnitude into the atom, replacing our spectrograph by a Fabry-Perot etalon, to find that even the fine structure levels are not levels after all: they are split by a yet smaller flea which is on their back to bite 'em. This is caused by two things: the fact that the nucleus is not a point-particle, and the fact that it can have a magnetic dipole moment.

First consider the charge distribution of the nucleus. The fact that it is spread over an approximately spherical volume of radius of order femtometres (10^{-15} m) introduces a shift in all the energy levels, which can be calculated very accurately using perturbation theory since it is small compared to the fine structure. The shift will be hard to determine experimentally in a given isotope, but it is noticeable because it will typically vary from one isotope to another of the same atom. The theoretical calculation for hydrogen is quite straightforward (for a graduate physicist at least—doable but perhaps not easy for an undergraduate!): we have to work out the difference between the electric field of a point particle and the electric field of a charged sphere, and treat this difference as a perturbation, performing the relevant integral to get $\delta E = \langle \delta H \rangle$.

Another contribution to the isotope shift was noted in section 2: the mass shift caused by the fact that the reduced mass of the electron depends on the mass of the nucleus.

Further effects cause not just shifts but also splittings of the levels within a given isotope.

13.1 Nuclear spin effect

The nuclear spin $\mathbf{I}$ interacts with the magnetic field at the nucleus due to the orbiting electrons:

$$H_{IJ} = -\boldsymbol{\mu}_I \cdot \mathbf{B}_{el}(0) = A\mathbf{I} \cdot \mathbf{J} \tag{130}$$

where the second version is appropriate since the effect is a small perturbation and the atomic states produce a net field along $\mathbf{J}$. Any magnetic field component along a direction other than $\mathbf{J}$ would be averaged to zero by all the precession going on inside the atom. This is the IJ coupling approximation, c.f. LS coupling, valid when the hyperfine structure is small compared to the fine structure, which is true for the great majority of neutral atoms (including all the cases you will meet).

Please notice that I have started to become lazy about putting hats on operators. This is ok because now we are good atomic physicists we can see at a glance where the operators are.

The magnetic moment of the nucleus is proportional to $\mathbf{I}$:

$$\boldsymbol{\mu}_I = g_I \mu_N \mathbf{I}. \tag{131}$$

Notice the absence of a minus sign in this expression, in contrast to (102). This is because the nucleus has a positive charge, so $\boldsymbol{\mu}_I$ usually points in the same direction as $\mathbf{I}$, but it doesn't have to: a nucleus can have a negative g_I. The proportionality constant is the *nuclear magneton* defined as

$$\mu_N = \frac{m_e}{m_p} \mu_B. \tag{132}$$

This definition is used since magnetic moments are typically inversely proportional to the mass of the particle, so this results in g_I values of order 1, but the g_I value does not have to be an integer (it is produced by complicated quark and nucleon interactions). For example the proton has $g_I = 5.585$. [Problem: the 'magnetic moment of the proton' is normally stated to be $2.79\mu_N$.

Is this consistent with what was just stated about g_I for the proton? What are the corresponding statements for the electron?]

Now that we have yet another angular momentum to add in, the truly total angular momentum of the atom becomes $\mathbf{F} = \mathbf{J} + \mathbf{I}$ and we obtain the exact expression $\mathbf{I} \cdot \mathbf{J} = (1/2)(\mathbf{F}^2 - \mathbf{J}^2 - \mathbf{I}^2)$ (cf eq. (116)). Hence the **hyperfine splitting** is

$$\langle IJF | H_{IJ} | IJF \rangle = \frac{1}{2} A \left(F(F+1) - J(J+1) - I(I+1) \right) \tag{133}$$

producing an interval rule, which will be broken if the electric quadrupole interaction (off syllabus, but explained below anyway) is non-zero.

To get the scaling of the hyperfine structure with Z for hydrogenic ions, think of the magnetic field produced by the electrons at the nucleus

fine:	magnetic field at electron from nucleus	nuclear charge Ze orbits with speed $-v$
hyperfine:	magnetic field at nucleus from electron	electron charge e orbits with speed v

conclusion: hyperfine structure scales as one power of Z less, i.e. Z^3, for hydrogen-like ions.

To be a bit more rigorous about it (but not much!), consider the magnetic field set up by two contributions. First we have the orbital motion of the electrons, and second the intrinsic magnetic dipoles of the electrons (due to their spin). The former gives a field of

$$-2\frac{\mu_0}{4\pi}\mu_B \left\langle \frac{1}{r^3} \right\rangle \hat{\mathbf{l}} \tag{134}$$

from the Biot-Savart law. A moment's thought reveals that the spin produces a contribution of the same order of magnitude (look up the magnetic field due to a magnetic dipole if you don't know it already).

We see that both contributions are important. The reason that the total field is nevertheless along $\mathbf{J}$ and not some other direction is that, in quantum language, we are considering eigenstates of $\hat{\mathbf{J}}^2$ and $\mathbf{J}_z$, or in semi-classical language the components along any other direction are averaged to zero by precession.

The case $l = 0$ is a special case (we met a similar special case when considering the fine structure of hydrogen) since (a) the electron isn't going around the nucleus and (b) $\langle 1/r^3 \rangle$ is not finite. The former means the orbit contribution to the field is zero, and in calculating the spin contribution we will find the $\langle 1/r^3 \rangle$ factor does not arise for $l = 0$. To calculate the magnetic field we consider the magnetic moment of the electron spread out over its wavefunction to produce a magnetization (magnetic dipole moment per unit volume) equal to $g_s \mu_B |\psi(r)|^2$. The field at the nucleus produced by this magnetization has zero net contribution from parts of the wavefunction outside the nucleus, because of the spherical symmetry. This is why no factor $\langle 1/r^3 \rangle$ arises. The part of the wavefunction inside the nucleus provides an essentially uniform magnetization, since $\psi(r) = \psi(0)$ to very good approximation for r of order the nuclear size. The field inside a uniformly magnetized sphere is

$$\mathbf{B} = \frac{2}{3}\mu_0 \mathbf{M} = -\frac{2}{3}\mu_0 g_s \mu_B |\psi(0)|^2 \mathbf{s}. \tag{135}$$

I don't think you need to know the detailed calculation for finals, but you may need to know that this is called the *Fermi contact interaction*. (The magnetic field inside a uniformaly magnetized sphere can be calculated in a few lines as follows: use the fact that the field outside the sphere is of dipole form with a dipole equal to the dipole of the sphere; remember the field due to a dipole; the normal component of $\mathbf{B}$ is continuous across the boundary, apply this along the axis of symmetry.)

Notice that by dimensional analysis, $|\psi(0)|^2$ is proportional to $1/a_0^3$, and for hydrogen-like systems we find

$$|\psi(0)|^2 = \frac{Z^3}{n^3 \pi a_0^3} \tag{136}$$

Putting this into equation (135) is a highly significant piece of physics, because it enables us to understand the 21 cm wavelength radiation produced by hydrogen atoms in their ground state. This is the wavelength of the transition between the two hyperfine levels of the ground state of hydrogen. Since the proton and the electron both have spin 1/2, we have $I = 1/2$, $J = 1/2$ giving $F = 1$ or 0. You are asked to calculate the splitting in the problems. The 21 cm radiation is observed across the sky, and enables us to map the distribution of hydrogen in the universe. The hydrogen maser is also one of the most stable oscillators we know how to build.

It so happens that the Fermi contact interaction is usually larger than the magnetic field produced by electrons of the same n but $l \neq 0$. A very rough way of understanding this is to argue
1. the field B_s from the electron **spin** is large for $l = 0$-states compared to $l > 0$ states because then the electron is close to the nucleus, and B_s varies as $1/r^3$.
2. when $l > 0$, the field B_l from the orbit is of the same order as B_s
therefore $B_s(l = 0) > B_{s \text{ and } l}(l > 0)$.

The size of hyperfine structure is about $m_e/m_p \simeq 0.0005$ times the size of fine structure. It is best measured in frequency units.

Another important example is the hyperfine splitting of the ground state of neutral caesium, because it is used to define our standard of time. Caesium is an alkali and so has $J = 1/2$ in the ground state. The nuclear spin is $I = 7/2$ leading to $F = 3$ or 4.

> The second is defined to be the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom.

This is discussed on the web-site of the UK National Physical Laboratory, which maintains our time and other standards, see http://www.npl.co.uk/, and use the site search engine to look for 'time'. The way the radiation period is observed precisely in the laboratory will be discussed in the final lecture; it is also discussed at the NPL site and at the US National Bureau of Standards, http://www.bldrdoc.gov/timefreq/cesium/fountain.htm.

The concept of a standard of time is worth thinking about a little. No one reports measurements of the value of the hyperfine splitting of caesium, because it is by definition equal to 9,192,631,770 Hz. Rather, if you have an oscillator in your laboratory which you think is oscillating at frequency f, then you should find that your oscillator produces f cycles during the time it takes for a caesium atom to go through 9,192,631,770 cycles. Therefore a very precise observation of caesium is needed to obtain a very precise knowledge of f. Measurements on caesium report the precision with which the frequency can be detected, and the stability from day to day of the apparatus (limited by varying fields and collisional effects, for example). Parts in 10^{-15} are quoted nowadays: a very nice discussion is by C. Salomon *at al.* in "Cold Atom Clocks," Atomic Physics 17: Proceedings of the XVII International Conference on Atomic Physics (ICAP 2000), pp 23-40 (American Institute of Physics, New York 2001).

Ultimately all statements about experimentally measured times, whether it be the orbital period of a planet, or the lifetime of pion, are statements about the ratio between the observed time and the period of the caesium clock.

13.2 Nuclear quadrupole term (off A-paper syllabus)

There is an electrostatic effect which arises from the departure of the shape of the nucleus from spherical symmetry. An isolated nucleus cannot carry a permanent electric dipole moment, so the first correction to the electrostatic potential is proportional to the *electric quadrupole moment* of the nucleus. One finds

$$H_Q = eQ \left\langle \frac{\partial^2 V_e}{\partial z^2} \right\rangle \frac{[3(\mathbf{I} \cdot \mathbf{J})^2 + \frac{3}{2}\mathbf{I} \cdot \mathbf{J} - I(I+1)J(J+1)]}{2I(2I-1)J(2J-1)} \tag{137}$$

where Q is the nuclear electric quadrupole moment operator, $-e\boldsymbol{\nabla} V_e$ is the electric field produced by the electrons at the nucleus, and I is the nuclear spin. The spin enters in this expression simply because the nuclear quadrupole moment must be aligned with the nuclear spin $\mathbf{I}$, so it provides a convenient way to identify the direction of the quadrupole moment. Note that $\langle \partial^2 V_e \partial z^2 \rangle \propto \langle 1/r^3 \rangle$.

13.3 Main points

The main points of this chapter have been:
isotope shift: mass shift (change in reduced mass), volume shift (finite volume of nuclear charge); hyperfine structure: chiefly magnetic effects, simple justification of formula $H = A\mathbf{I} \cdot \mathbf{J}$; Z^3 scaling in hydrogen-like ions; relative size of hyperfine compared to fine structure $((1/Z)(m_e/m_p))$; fact that s-states ($l = 0$) have large hyperfine splitting; Fermi contact interaction; 21 cm transition in hydrogen; time standard is hyperfine structure of ground state of caesium.

14 The Zeeman effect

We have now finished our discussion of a free atom or ion, that is, one with no external influences on it. [Problem (hard): is it possible that there is substructure within the hyperfine levels of an atom?] We now turn to the influence of external things such as magnetic and electric fields, and electromagnetic radiation.

When an atom is placed in a uniform magnetic field B, the transitions are observed to split into several components. This is called the Zeeman effect. When the field is large (exactly how large will be seen below) the way the components vary with B changes and this regime is called the Paschen-Back effect.

14.1 Weak B field

We will calculate the weak-field Zeeman effect as if we are treating LS coupling, but the treatment will be completely general. The result for any other angular momentum coupling $j_1 j_2$ is obtained by substituting j_1 for L and j_2 for S, and putting in the correct g-factors. We will look at how this works in the case of hyperfine structure in the next section.

An external magnetic field will interact primarily with the net magnetic dipole moment $\boldsymbol{\mu}$ of an atom if it has one. [Problem: is it possible for an atom to have zero magnetic dipole moment?] The interaction Hamiltonian is therefore

$$H = -\boldsymbol{\mu} \cdot \mathbf{B} \tag{138}$$

14.1.1 Normal Zeeman effect

First consider the case of an LS coupled atom in a singlet state, $S = 0$. Then

$$\boldsymbol{\mu} = -g_l \mu_B \hat{\mathbf{L}}. \tag{139}$$

For $S = 0$ the total magnetic moment is $\mathbf{J} = \mathbf{L}$. Before the field is applied $\mathbf{J}$ and therefore here $\mathbf{L}$ is a constant of the motion. Choosing the z axis to lie along $\mathbf{B}$, we obtain

$$-\boldsymbol{\mu} \cdot \mathbf{B} = -\mu_z B = g_l \mu_B B \hat{L}_z \tag{140}$$

and the effect is

$$\Delta E = \langle LM_L | \, g_l \mu_B B \hat{L}_z \, | LM_L \rangle = g_l \mu_B B M_L. \tag{141}$$

The $\mathbf{L}$ vector which was fixed in space is now precessing about $\mathbf{B}$, but clearly $\hat{L}_z$ is still a constant of the motion.

The energy level L is now split into $2L + 1$ levels (corresponding to the different values of M_L) separated by $\mu_B B$ (since $g_l = 1$).

A radiative transition involves transitions between two different states. If these are both singlets, then both will have the same Zeeman splitting $\mu_B B$ and the resulting pattern of emitted frequencies is called the Normal Zeeman effect. Owing to conservation of angular momentum, the atom's angular momentum projection on the z axis can change by at most $\hbar$ when it emits or absorbs a photon by electric dipole radiation (see also section 15). The three possible types of transition $M_L' = M_L + 1$, $M_L' = M_L$ and $M_L' = M_L - 1$ have frequencies shifted with respect to the zero-field frequency by $+\mu_B B$, 0, and $-\mu_B B$ respectively.

14.1.2 Zeeman effect

The fact that g_s is not equal to 1 means that the Zeeman effect is much richer when $S \neq 0$. Because the classical model of angular momentum only allows $g = 1$, this more general case was originally called 'anomalous', although it is in fact the more typical case, so we tend to drop the adjective 'anomalous' nowadays in the laboratory, though the word lives on in textbooks and finals questions.

It is rather nice to see what happens when both $\mathbf{L}$ and $\mathbf{S}$ are involved. The magnetic dipole moment operator for an LS coupled atom of total angular momentum $\mathbf{J}$ is

$$\boldsymbol{\mu} = -\left(g_l \mu_B \mathbf{L} + g_s \mu_B \mathbf{S}\right). \tag{142}$$

Note that since $g_s \neq g_l$, $\boldsymbol{\mu}$ is not parallel to $\mathbf{J}$. The Zeeman interaction Hamiltonian is

$$H = -\boldsymbol{\mu} \cdot \mathbf{B} = g_l \mu_B B \hat{L}_z + g_s \mu_B B \hat{S}_z \tag{143}$$

where we have chosen the z axis to be along the $\mathbf{B}$ field (we are always free to do this).

You might expect that the energy shift is given by

$$\delta E \stackrel{?}{=} g_l \mu_B B M_L + g_s \mu_B B M_S \tag{144}$$

but we have to remember that in LS coupling M_L and M_S are not good quantum numbers. $\mathbf{L}$ and $\mathbf{S}$ are precessing about $\mathbf{J}$ in the vector model, and since $\mathbf{J}$ is not along $\mathbf{B}$ the z-components of $\mathbf{L}$ and $\mathbf{S}$ are not constants of the motion. All this is simply lots of ways of saying that the energy

eigenstates of the atom, before the perturbation is applied, are not eigenstates of L_z or S_z; they are eigenstates of $\mathbf{L}^2$, $\mathbf{S}^2$, $\mathbf{J}^2$ and J_z.

In first order perturbation theory we are calculating, therefore,

$$\delta E = \langle LSJM_J | \, g_l\mu_B B\hat{L}_z + g_s\mu_B B\hat{S}_z \, | LSJM_J \rangle . \tag{145}$$

Taking the $\hat{L}_z$ part first, use the following observation from the vector model: $\mathbf{L}$ is precessing rapidly about $\mathbf{J}$, so the mean value of L_z is found by first projecting $\mathbf{L}$ onto $\mathbf{J}$, and then taking the component of this vector along the z-direction. The expressions for the classical vectors are as follows: the size of the component of $\mathbf{L}$ along $\mathbf{J}$ is $\mathbf{L} \cdot \mathbf{J}/J$. The direction of this projected vector is along $\mathbf{J}$, so to get its z component we need to multiply its length by J_z/J. Hence when $\mathbf{L}$ precesses about $\mathbf{J}$ the mean value of L_z is $(\mathbf{L} \cdot \mathbf{J})J_z/J^2$.

The exact result for the quantum operators is:

$$\langle LSJM_J | \, \hat{L}_z \, | LSJM_J \rangle = \frac{\langle LSJM_J | \hat{\mathbf{L}} \cdot \hat{\mathbf{J}} | LSJM_J \rangle \langle LSJM_J | \, \hat{J}_z \, | LSJM_J \rangle}{\langle LSJM_J | \hat{\mathbf{J}} \cdot \hat{\mathbf{J}} | LSJM_J \rangle} \tag{146}$$

—the classical vector model serves simply to show that this expression is reasonable, but at this level we have to quote the quantum operator expression without any more complete proof[13]. In order to bring out the form of this result, I will write it again without the state labels, but one should be aware nonetheless of what type of angular momentum state we are considering:

$$\left\langle \hat{L}_z \right\rangle = \frac{\left\langle \hat{\mathbf{L}} \cdot \hat{\mathbf{J}} \right\rangle \left\langle \hat{J}_z \right\rangle}{\left\langle \hat{\mathbf{J}} \cdot \hat{\mathbf{J}} \right\rangle} \qquad [\text{ for } |LSJM_J\rangle \tag{147}$$

All the expectation values here are trivial to evaluate in the state under consideration:

$$\langle LSJM_J | \, \hat{J}_z \, | LSJM_J \rangle = M_J \tag{148}$$

$$\langle LSJM_J | \, \hat{\mathbf{J}}^2 \, | LSJM_J \rangle = J(J+1) \tag{149}$$

$$\langle LSJM_J | \, \hat{\mathbf{L}} \cdot \hat{\mathbf{J}} \, | LSJM_J \rangle = \frac{1}{2} \left(J(J+1) + L(L+1) - S(S+1) \right) \tag{150}$$

where for the last result we used

$$\begin{aligned} \mathbf{J} &= \mathbf{L} + \mathbf{S} \\ \implies (\mathbf{J} - \mathbf{L})^2 &= \mathbf{S}^2 \\ \implies \mathbf{J}^2 + \mathbf{L}^2 - 2\mathbf{L} \cdot \mathbf{J} &= \mathbf{S}^2 \\ \implies \mathbf{L} \cdot \mathbf{J} &= \frac{\mathbf{J}^2 + \mathbf{L}^2 - \mathbf{S}^2}{2} \end{aligned}$$

which is an exact relationship between the operators. Therefore we have

$$g_l\mu_B B \left\langle \hat{L}_z \right\rangle = g_l\mu_B B \frac{\left\langle \hat{\mathbf{J}}^2 + \hat{\mathbf{L}}^2 - \hat{\mathbf{S}}^2 \right\rangle}{2J(J+1)} M_J. \tag{151}$$

The treatment of the $\hat{S}_z$ part in (145) goes on exactly corresponding lines; it is easy to see that the result is as before except that L and S are swapped. Since the overall result is proportional to M_J, we define a Landé factor g_J for the total J by the expression

$$\Delta E = \langle -\boldsymbol{\mu} \cdot \mathbf{B} \rangle = g_J\mu_B B M_J, \qquad [\text{definition of } g_J \tag{152}$$

[13]The standard proof invokes the Wigner Ekart theorem, it is outlined in Woodgate.

and putting it all together we have

$$g_J = g_L \frac{J(J+1) + L(L+1) - S(S+1)}{2J(J+1)} + g_S \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \tag{153}$$

If we put $g_L = 1$ and $g_S = 2$ this expression can be simplified somewhat:

$$g_J = \frac{3J(J+1) - L(L+1) + S(S+1)}{2J(J+1)}. \tag{154}$$

It is useful to notice the following simple cases:

when $S = 0$, $g = 1$ (since purely orbital)
when $L = 0$, $g = 2$ (since purely spin)
when $S = L$, $g = 3/2$

It is also worth noticing that our derivation was quite general to obtain the net g factor when two angular momenta of known g factors are coupled together. For example, we can see immediately that the Zeeman effect of a hyperfine structure level in IJ coupling is $g_F \mu_B M_F$ with

$$g_F = g_J \frac{F(F+1) + J(J+1) - I(I+1)}{2F(F+1)} - g_I \frac{m_e}{m_p} \frac{F(F+1) + I(I+1) - J(J+1)}{2F(F+1)}, \tag{155}$$

where the factor m_e/m_p allows for the fact that g_I is defined with a scale factor of μ_N rather than μ_B, and the minus sign comes from the different sign convention in the definition of g_I compared with g_J.

You should derive g_J and examine carefully the Zeeman splittings for some example cases such as the transitions $^2S_{1/2} \to\, ^2P_{3/2}$ and $^3P_1 \to\, ^3P_2$. Make sure you understand not only the energy level splittings, but also how all the transition frequencies in the Zeeman spectrum arise, bearing in mind that only $\Delta M_J = \pm 1, 0$ transitions arise.

14.1.3 Directional dependence and polarization of the emitted light

When you look at the light emitted along the direction of the magnetic field $\mathbf{B}$, you don't see all the components of the Zeeman spectrum! The ones corresponding to $\Delta M = 0$, the so-called π components, are not there. Yet they are there when you look in any direction perpendicular to the magnetic field lines (together with the $\Delta M = \pm 1$ components which are called $\sigma_\pm$). What is going on? …In a $\Delta M = 0$ transition, the atomic charge distribution is oscillating along the magnetic field direction, and therefore the emitted light is polarized parallel to $\mathbf{B}$. But light is a transverse oscillation: light can't propagate in the direction along which its electric field is oscillating. The intensity of the π light therefore falls gradually to zero as a function of angle out of the plane perpendicular to $\mathbf{B}$.

Meanwhile the $\Delta M = \pm 1$ transitions are associated with rotation of the atomic charge distribution in the plane perpendicular to $\mathbf{B}$. The light comes out in all directions but oscillates always in this plane. Looking along the plane, therefore, we see linear polarization in the plane, i.e. perpendicular to the π polarization. Looking perpendicular to the plane, i.e. along $\mathbf{B}$, we see circular polarization. That is, the $\Delta M = +1$ transition gives right-handed circular polarization, $\Delta M = -1$ gives left-handed circular polarization. [Problem: if you look at the Zeeman spectrum in a direction perpendicular to $\mathbf{B}$, and slowly rotate a linear polarizer in front of your eye, what do you expect to observe?]

Uses of the Zeeman effect

With modern computers we can make a fair stab at calculating atomic structure from scratch, but the problem is so complicated for some atoms that even this is only moderately successful. Historically, the problem of working out the identities of all the different atomic energy levels from the observed spectral lines was like trying to deduce the structure of a piano from the noise it makes when thrown down stairs. The Zeeman effect is a very powerful tool in this analysis. Merely by counting components in the Zeeman spectrum one can make a pretty fair guess at the values of J for the two levels involved in the transition. The relative positions of the Zeeman lines then lead rapidly to confirming J and identifying L and S, or indeed to concluding that LS coupling is not a good approximation. One of the problems gives you some practice at such a piece of detective work.

Nowadays, the mere identities of atomic levels are not exactly hot news, but occasionally we need to know the details of some particular transition, in order to use it for some purpose, and the vast majority of transitions in all but the most simple atoms will (rightly) forever remain uncalculated and unobserved until they are needed. The Zeeman effect therefore retains its old usefulness as an identifier of levels.

However, the more common uses of the Zeeman effect are as a diagnostic of magnetic fields in an inaccessible environment such as the atmosphere of a star or the inside of a tokamak fusion reactor, and as a tool for manipulating atoms. The Zeeman effect is used for example to define an axis in space so that atoms in atomic clocks don't wander in direction. It is also a central ingredient in other magnetic resonance experiments. A more ingenious use is in atom traps, where a magnetic field gradient provides a force to confine cold atoms. A Zeeman effect which varies with position is also used in one laser cooling method to compensate for the changing Doppler effect when atoms are slowed down to very low speeds by laser light.

14.2   Strong B field: Paschen-Back effect

The above perturbation theory is valid as long as the Zeeman shift ΔE is small compared to the separation of the energy eigenstates before the perturbation was applied. This is typically ok for fine structure splittings ($\sim$ 100 GHz so $B \ll 7$ Tesla is 'small'), but not always for hyperfine structure ($\sim$ 100 MHz so $B \ll 0.007$ Tesla required for 'small').

If the Zeeman effect dominates the fine structure, then we have to apply the perturbations in reverse order, i.e. first Zeeman, then spin-orbit. The result for Zeeman splitting is then as given in equation (144) since now we can argue that M_L and M_S are well-defined. This situation is called the Paschen-Back effect. The angular momentum vectors $\mathbf{L}$ and $\mathbf{S}$ now precess independently about the magnetic field direction. In the Pashchen-Back situation, spin-orbit coupling produces a further (comparitively small) shift proportional to $\langle LSM_LM_S| \mathbf{L} \cdot \mathbf{S} |LSM_LM_S\rangle = M_LM_S$.

In going from weak to strong field, the good quantum numbers therefore change from L, S, J, M_J to L, S, M_L, M_S: note that the net angular momentum along the z direction is a constant of the motion at all field strengths, this can be underlined by referring to a quantum number M which is equal to M_J at small B and to $M_L + M_S$ at large B.

Similar arguments apply to the case of the Zeeman effect in hyperfine structure. The case where $J = 1/2$ can be solved exactly, leading to the Briet-Rabi formula, valid for all field strengths:

$$\Delta E(F, M_F) \quad = \quad g_I\mu_B B M_F + \frac{\Delta E_0}{2}\left(\frac{-1}{2I+1} \pm \sqrt{1 + \frac{4M_F}{2I+1}x + x^2}\right) \tag{156}$$

$$\text{where} \quad x \quad = \quad (g_J - g_I')\mu_B B/\Delta E_0 \tag{157}$$

You might like to amuse yourself by plotting this formula by computer (or by hand) so as to see the behaviour of the hyperfine levels in hydrogen, for example, which are crucial in the hydrogen maser (see section 16).

14.3   External electric field: the Stark effect (off syllabus)

A uniform external electric field $\mathbf{E}$ will also perturb an atom. The details are somewhat different since the atom does not have an electric dipole moment before the field is applied: rather, the field polarizes the atom. See textbooks for details.

The oscillating electric field of a light wave produces an oscillating electric dipole moment in an atom, and an associated shift called the a.c. Stark shift, or light shift. This is a versatile tool for laser-manipulation of atoms (both cooling and trapping), but is beyond the level of these notes.

14.4   Main points

$H = -\boldsymbol{\mu} \cdot \mathbf{B}$; normal Zeeman effect giving three transition lines; anomalous Zeeman effect: vector model description; derivation of g factor in LS coupling; how the transitions divide up into several components; polarization of the light; Paschen-Back effect in strong fields; uses of the Zeeman effect: identifying levels, probing magnetic fields, manipulating atoms.

15 Atom–light interaction

One of the most immediately obvious and important features of atoms is that they interact strongly with light. A single sodium atom illuminated by a modest laser beam will scatter 30 million photons per second. This is a lot of photons for such a small emitter! A human dark-adapted eye looking at this atom from a distance of 10 cm would receive around 6000 photons per second. [Problem: check this calculation.] This is enough to make the light emitted by a single atom detectable to the naked eye.

The primary way in which electromagnetic waves are produced is by **accelerating charge**. (Recall Brehmsstralung or a simple radio antenna for example). Before quantum mechanics was understood, it was thought that the electrons in an atom could not be orbiting the nucleus because if they were then the high accelerations involved would produce strong electromagnetic radiation and the electrons would rapidly lose energy and plunge into the nucleus in a fraction of a nanosecond. There is some correct insight in this argument: any accelerating charge will indeed emit electromagnetic waves, quantum mechanics is fully in agreement with that. The reason that the energy eigenstates in an atom are much more stable than the classical argument suggests is that the charge is not accelerating: it is smeared out around the atom and *does not move* when the atom is in an energy eigenstate. To be precise, the probability distribution $|\psi|^2$ of the electrons is completely independent of time. This is in fact just the standard property of energy eigenstates which you should know already: it is why they are called stationary states.

Electromagnetic radiation happens when an atom is in a **superposition of energy eigenstates**. Consider two energy eigenstates, for example ψ_{1s} and ψ_{1p} having energies E_1 and E_2 respectively. According to the time-dependent Schrödinger equation, the wavefunction of each of these states has a time-dependence $\exp(-iEt/\hbar)$. Therefore if the atom is in a superposition of the two, then its wavefunction depends on time in the following way:

$$\psi_{\mathbf{r},t} = c_1\psi_{1s}(\mathbf{r})e^{-iE_1t/\hbar} + c_2\psi_{2p}(\mathbf{r})e^{-iE_2t/\hbar} \tag{158}$$

The charge density of the electron is equal to

$$-e|\psi|^2 = -e\left(|c_1\psi_{1s}|^2 + |c_2\psi_{2p}|^2 + 2\mathrm{Re}\left[c_1^*c_2\psi_{1s}^*\psi_{2p}e^{-(E_2-E_1)t/\hbar}\right]\right) \tag{159}$$

To be absolutely clear on what is going on, we will take the case that $c_1\psi_{1s}$ and $c_2\psi_{2p}$ are real, though the argument will not depend on this. Then we have

$$-e|\psi|^2 = -e\left(c_1^2\psi_{1s}^2 + c_2^2\psi_{2p}^2 + 2c_1c_2\psi_{1s}\psi_{2p}\cos\omega_{21}t\right) \tag{160}$$

where $\omega_{21} = (E_2 - E_1)/\hbar$. We now have a charge density distribution which is oscillating at the frequency ω_{21}. That is how an atom can emit light, and that is why the frequency of the light waves coming out is equal to the frequency difference of two energy eigenstates of the atom.

Once the atom is in such a superposition, therefore, light waves[14] are emitted. Therefore the atom must be losing energy. In terms of the wavefunction, what happens is that the coefficients c_1 and c_2 vary (smoothly) in time: c_2 decreases and c_1 increases. Typical timescales for a strong optical transition at 600 nm wavelength are as follows:

period of one oscillation of the emitted light	$2\text{ fs} = 2 \times 10^{-15}$ s
time for c_2 to fall to 0 and c_1 to rise to 1	$20\text{ ns} = 2 \times 10^{-8}$ s
$\Rightarrow$ number of oscillation periods in the emitted pulse	10^7

The final number is also the number of wavelengths long which the pulse of light is; we find a length of 6 metres in this example (but note the pulse length will be much different for other transitions).

---

[14]Throughout this section when I refer to light waves, of course in general I mean electromagnetic waves of frequency ω, which might be anywhere from radio waves to X rays etc.

There are two things which remain to be worked out: first, how does the atom get into a superposition of energy eigenstates in the first place, and secondly what determines the speed with which the coefficients c_1 and c_2 change. More fully, we would like to know how the coefficients change as a function of time: you should not be surprised to learn that the answer is exponential decay of $|c_2|^2$, and $|c_1|^2 = 1 - |c_2|^2$.

The atom gets into a superposition of energy eigenstates because it is jiggled by a light wave of the right frequency. This is called stimulation. However, it is remarkable that even in the absence of any light, atoms still decay. We now understand this in terms of the quantum vacuum state: the vacuum state has zero mean energy, but its r.m.s. energy is not zero: there are continual fluctuations at all frequencies, and these stimulate atoms to decay.

The calculation of the atomic decay in terms of these vacuum fluctuations is very complicated, and beyond undergraduate level. However, there is a very beautiful arguement due to Einstein which links the spontaneous decay process to two other processes, namely absorption and stimulated emission. The latter can both be understood in terms of standard quantum mechanics, section 15.4, so Einstein's argument plus standard quantum mechanics gives us everything we need.

15.1   Einstein A and B coefficients

Einstein's argument really is rather wonderful. It uses thermodynamics to show that the light-emitting and -absorbing properties of atoms must obey a certain pair of relationships. It is remarkable that quantum theory is constructed in such a way that it obeys the principles of thermodynamics. This is an illustration of some very deep connections between our basic physical theories.

In the argument, we appeal to the case of thermal equilibrium in order to derive two relationships between three parameters A_{21}, B_{21} and B_{12}. But note, once we have derived the relationships, they hold true **whether or not the atom or the radiation is in thermal equilibrium.** This is analogous to the argument leading to the Kirchhoff theorem in the thermodynamics of radiation and matter: the fact that good emitters are good absorbers is true in everyday situations as well as in a thermal cavity.

Consider a collection of atoms with two energy levels 1 and 2 having degeneracies g_1 and g_2 (level 1 is the lower state, level 2 the upper state). The *total* population of the lower level (i.e. the number of atoms in that level) is N_1, and the population of level 2 is N_2. (With g_1 states per level, the population of each state is N_1/g_1, and similarly for N_2/g_2.)

There are three radiation processes: spontaneous emission, absorption, and stimulated emission (see diagram in the lecture). In the presence of radiation we **assume** that the absorption and stimulated emission rates will be proportional to the **energy density per unit frequency range** in the radiation, and we **assume** that the spontaneous emission rate depends only on the atom, not on the radiation at all. Those are the only assumptions made. They are very reasonable, though the first is maybe a bit new if you aren't too familiar with classical electromagnetic radiation physics.

To show that spectral energy density is the relevent quantity for absorption: We take it as reasonable that the rate of atomic transitions of the absorption type should be proportional to the energy of the radiation. By a collision-type argument, the atom will absorb some fraction of the light crossing unit area per unit time (the light energy *flux*), and energy flux F is proportional to energy density, i.e. energy per unit volume, u. [Problem (easy): prove that $F = uc$ for a plane wave of light]. So far we have absorption rate proportional to u. Now, suppose the atom is hydrogen and

all the radiation is in the wavelength range 200 nm to 600 nm. But hydrogen doesn't have any transitions in this wavelength range! We would surely expect hydrogen to absorb at a much higher rate if the radiation had the *same* energy per unit volume but was concentrated in wavelength around 121.57 nm (the $n = 1 \rightarrow 2$ transition in hydrogen). This shows that the relevant quantity is not the energy density $u(\omega)$, but the energy density per unit frequency range $\rho(\omega) = u(\omega)/\Delta\omega$ in the vicinity of $\omega_{21} = (E_2 - E_1)/\hbar$, for a transition between energy levels E_1 and E_2.

With the assumptions given above, we must have three *coefficients* which depend only on the atom, not the radiation. These are A_{21}, B_{21} and B_{12}. The easiest way to define them precisely is simply to write the formulae for the net transition rates into levels 1 and 2, which must be, under the assumptions given,

$$\frac{dN_1}{dt} = (B_{21}N_2 - B_{12}N_1)\rho(\omega_{21}) + A_{21}N_2 \tag{161}$$

$$\frac{dN_2}{dt} = -\frac{dN_1}{dt}, \tag{162}$$

where $\rho(\omega)$ is the energy density per unit frequency interval in the incident light. Thus B_{21} is the rate coefficient of transitions from 2 to 1 (i.e. stimulated emission) per unit spectral energy density, B_{12} is the rate coefficient of transitions from 1 to 2 (i.e. absorption) per unit spectral energy density, and A_{21} is the rate coefficient of spontaneous transitions from 2 to 1. (There are no spontaneous transitions upwards in energy from 1 to 2).

This defines the Einstein coefficients A_{21}, B_{21}, B_{12}. If we now assume the special case of an atom in thermal equilibrium with thermal radiation at temperature T, then we can make use of three standard results from the theory of thermal equilibrium:

$$\dot{N}_1 = \dot{N}_2 = 0, \quad \text{[steady state} \tag{163}$$

$$\frac{N_2}{N_1} = \frac{g_2}{g_1}e^{-\hbar\omega_{21}/k_BT}, \quad \text{[Boltzmann distribution} \tag{164}$$

$$\rho(\omega_{21}) = \frac{\hbar\omega_{21}^3}{\pi^2 c^3}\frac{1}{e^{\hbar\omega_{21}/k_BT} - 1}. \quad \text{[Black body radiation} \tag{165}$$

However, at $\dot{N}_1 = 0$, eq. (161) implies

$$\rho(\omega_{21}) = \frac{A_{21}}{B_{12}N_1/N_2 - B_{21}} \tag{166}$$

so we have a consistent solution for all values of T if and only if

$$g_1 B_{12} = g_2 B_{21}, \quad \text{and} \quad A_{21} = \frac{\hbar\omega_{21}^3}{\pi^2 c^3}B_{21} \tag{167}$$

This is the only solution because A_{21}, B_{21} and B_{12} are properties of the atomic transition, so they do not themselves depend on temperature.

That's it! We have proved that in conditions of thermal equilibrium, relations (167) hold. But by assumption, the coefficients A_{21}, B_{21} and B_{12} are intrinsic properties of the atomic transition under consideration, they do not depend for example on the relative populations N_2 and N_1, so they don't depend on whether the atom is in thermal equilibrium or not. Neither do the *coefficients* depend on the radiation at all (think carefully about that statement). We needed black body radiation for the *derivation*, but the result is then simply true, no matter how we derived it.

The distinction between which statements are true only in thermal equilibrium, and which are true both in and out of thermal equilibrium, is one which comes up in many thermodynamic arguments. **Students usually get confused about this!** In the present argument, eq.s (163) to (165) are only valid for thermal equilibrium, but equations (161), (162), and (167) are always valid.

Most situations which we consider in atomic physics are **not** in thermal equilibrium. For example in an absorption experiment the gas sample may initially be in thermal equilibrium, but as soon as we shine non-black-body radiation at it, the populations of the energy levels will change, and they will no longer be given by the Boltzmann distribution. In this situation the atoms of the gas may still have their *motion* in thermal equilibrium at room temperature, but the *internal* states of the atoms may be far from equilibrium, and not described by any temperature.

15.1.1   Comments

There are several observations worth making from (167). First, when $g_1 = g_2$ we have $B_{21} = B_{12}$. That means electromagnetic radiation has *exactly* the same propensity to cause stimulated emission as absorption. Which process occurs the most will depend on the populations N_1 and N_2—see the discussion of the laser in the next section. In fact, although we have not proved it, in stimulated emission the emitted light has exactly the same frequency, direction and polarization as the incident light, and is even emitted exactly in step with (in phase with) the oscillations of the incident wave. We say it is emitted into the same *mode* of the radiation field. This is in contrast to spontaneous emission, where the photon is emitted in a random direction and polarization, and within a spread of frequencies near ω_{21}—see section 15.3

It is noteworthy that the ratio of total emission rate to absorption rate is

$$\frac{P(\text{spontaneous emission}) + P(\text{stimulated emission})}{P(\text{absorption})} = \frac{1 + \bar{n}}{\bar{n}} \qquad (168)$$

where $\bar{n}$ is the mean number of photons per mode of the incident radiation field. [Problem: prove this for the case of Black Body radiation by employing (161) and (167) and using statistical mechanics to obtain $\bar{n}$.] This hints at the full quantum description of light in which fields are like harmonic oscillators, and n counts the harmonic oscillator energy levels of energy $(n + 1/2)\hbar\omega$.

With no incident radiation present, we have

$$\frac{dN_2}{dt} = -A_{21}N_2, \qquad (169)$$

therefore N_2 decays exponentially with time-constant $\tau = 1/A_{21}$ (this is the variation of $|c_2|^2$ noted in the introductory discussion). A large A_{21} means a short lifetime of the excited state. Observe from eq. (167) that two transitions with the same B coefficient but of different frequency have widely differing A coefficient because of the ω^3 dependence. For example, an optical transition at 600 nm will give an A coefficient larger than that of a microwave transition at 1 GHz (of the same B_{21}) by the huge factor $(5 \times 10^{14}/1 \times 10^9)^3 = 1.25 \times 10^{17}$. As a consequence, microwave transitions are so slow as to almost never occur spontaneously, while X-ray transitions are extremely fast.

15.2   The laser

In this section we will make the simplifying assumption $g_2 = g_1$ (same number of states in both levels, e.g. a transition which does not involve a change in J). It is left as an exercise for the reader to generalize to $g_2 \neq g_1$.

For a normal gas of atoms in thermal equilibrium, a beam of light passing through the gas will be absorbed. But the gas doesn't get hotter and hotter—where has the light energy gone? What really happens is that absorption removes photons from the light beam, stimulated emission puts some back, and the rest of the energy is reradiated as spontaneous emission which goes in all directions. Assuming a steady-state situation has been reached, we have $dN_1/dt = 0$ and therefore equation (161) (which is true in all situations) tells us

$$\frac{N_2}{N_1} = \frac{B_{12}\rho}{B_{21}\rho + A_{21}}. \tag{170}$$

As ρ increases, N_2/N_1 gets larger, but it never rises above 1 (for $g_2 = g_1$). This means that the atoms are at most half excited, half not. Therefore the rate of stimulated emission $N_2 B_{21}\rho$ never exceeds the rate of absorption $N_1 B_{21}\rho$. The light is always absorbed on average, and its intensity falls exponentially with distance as it propagates through the gas.

If we could think of a way to achieve $N_2 > N_1$ the converse would be true: the stimulated emission would exceed the absorption, and the light intensity would grow exponentially as the light propagates through the gas. This is what happens inside a laser. The condition $N_2 > N_1$ is called *population inversion*. Population inversion is most easily achieved for a pair of levels where the lower level is not the ground state of the atom, so it too can decay. One good scheme is to identify an atom with four levels $0, 1, 2, 3$ which have the properties:
1. Level 0 does not decay (it is the ground state)
2. Transitions $0 \to 3$ can be excited by electron bombardment in a discharge.
3. Spontaneous transitions from level 3 to level 2 are significantly more likely than 3 to 1.
4. Level 2 has a longer natural lifetime than level 1.

[Problem: (**very** hard except by looking up known cases!): can you find four levels in some atom or ion which approximate these conditions?] You should be able to convince yourself that these conditions will result in a population inversion between levels 2 and 1, so 'optical gain' and hence laser light will be produced on the 1–2 transition. A more complete analysis would involve solving the *rate equations* for this situation. You should be able to see that they are:

$$\begin{aligned}
\frac{dN_3}{dt} &= R - A_{32} N_3 \\
\frac{dN_2}{dt} &= A_{32} N_3 - (B_{21} N_2 - B_{12} N_1)\rho(\omega_{21}) - A_{21} N_2 \\
\frac{dN_1}{dt} &= (B_{21} N_2 - B_{12} N_1)\rho(\omega_{21}) + A_{21} N_2 - A_{10} N_1
\end{aligned}$$

where R is the rate of excitation of 3 by the discharge (we have assumed this rate is quite unrelated to the population N_3 which is reasonable). We have also assumed that 3 decays only to 2, and we have neglected the effect of the very weak radiation from this spontaneous transition and from the spontaneous transition from 1 to 0. [Problem (not hard): show that these rate equations allow $N_2 > N_1$ in steady state provided $A_{10} > A_{21}$.]

More generally, it is not hard to show that a steady-state population inversion can only be maintained between two levels for which the upper has a longer lifetime than the lower.

When $N_2 > N_1$ any small spontaneous emission on the transition 2 to 1 will start a light beam which is then rapidly amplified as it propagates. The defining characteristics of laser radiation compared to light from other sources are the combination of brightness, directionality and extremely high monochromaticity. All these properties arise from the fact that the mean number of photons per mode $\bar{n} \gg 1$ in laser light, whereas $\bar{n} < 1$ in other light (e.g. $\bar{n} \simeq 0.06$ for the brightest spectral region of black body radiation at temperature 1000 K).

Note that we do not require mirrors to make a laser. However, a pair of highly reflecting mirrors is useful to recirculate the light and allow it to be amplified many times. One mirror has a small transmissivity $T \simeq 0.01$ to let the light get out. With $T = 0.01$ the light inside the laser is 100 times brighter even than the light which comes out! We need a small T because typically the net fractional increase in light intensity given by one to-and-fro pass through the gas (the so called *round-trip gain*)is only a few percent. We need the round trip gain to exceed the round trip loss $T + a$ where a accounts for absorption in the mirrors and other elements.

15.3   Natural linewidth and shape; Doppler broadening

During *spontaneous* emission, N_2 decreases exponentially with time, and this means that the pulse of light emitted decreases exponentially in intensity as $\exp(-A_{21}t)$, therefore in amplitude as $\exp(-A_{21}t/2)$. Now, any wave which is finite in duration has a non-zero frequency spread. The frequency analysis of the emitted wave is obtained by Fourier analyzing $\exp(-A_{21}t/2)\cos(\omega_{21}t)$; the integral is quite straightforward. Taking the modulus squared of the result we obtain for the intensity of each frequency component of the emitted light:

$$I(\omega) \propto \frac{1}{(\omega - \omega_{21})^2 + A_{21}^2/4}. \tag{171}$$

This distribution is called a Lorentzian distribution, it was plotted in the lecture. You should learn as a matter of general knowledge that the Fourier transform of an exponential function is a Lorentzian function. [Problem (easy): show that the full width of this curve, at half its maximum height (the FWHM) is equal to A_{21}].

Therefore A_{21} gives the *linewidth* of the spontaneously emitted radiation, in angular frequency units. In cyclic frequency units, the linewidth is $A_{21}/2\pi$.

The light emitted by a collection of atoms at room temperature will typically have a larger spread in frequencies than this natural linewidth. The major contribution for optical transitions in atoms at room temperature is the *Doppler broadening*. Each atom emits a fairly well-defined frequency $f = f_{21} \pm A_{21}/2\pi$ in its own reference frame, but owing to the motion of the atom with respect to our measuring instrument, the frequency observed is Doppler-shifted. A gas of atoms has a spread of velocities and therefore a spread of Doppler shifts. If this spread is large compared to $A_{21}/2\pi$ then the observed radiation will have a distribution in frequency dominated by this *Doppler broadening*.

For velocities v small compared to the speed of light[15], the Doppler shift δf is given by

$$\frac{\delta f}{f} = \frac{v}{c}. \tag{172}$$

Kinetic theory tells us that the velocity component along any one direction has a spread given by $(1/2)m(\Delta v)^2 = (1/2)k_B T$ where m is the mass of the atom. Therefore the frequency spread from all the different Doppler shifts is

$$\Delta f = f\frac{\Delta v}{c} = \frac{1}{\lambda}\sqrt{\frac{k_B T}{m}} \tag{173}$$

The frequency distribution exactly maps the velocity distribution, and is therefore Gaussian in shape. [Problem: why do we only consider one component of the velocity?]

---

[15]Don't forget we are thinking about the translational motion of the whole atom here, not the vibrational motion of the electron cloud around the atom.

For optical transitions at room temperature, typical Doppler broadening is of order some hundreds of MHz, so it typically blurs away hyperfine structure of the spectral lines, but not fine structure, except in hydrogen. Doppler-broadening is close to a non-issue when observing microwave radiation, on the other hand, because the Doppler spread is proportional to the radiation frequency. [Problem: calculate Δf due to Doppler broadening for sodium resonance lines at 100 °C. Compare your answer with the natural linewidth of 10 MHz.] [Problem: The fine structure splitting of the $n = 3$ level in hydrogen is almost completely washed out when observations are made of the $2s \rightarrow 3p$ transition at room temperature. However, the same fine structure is very well resolved when obervations are made of the $3s \rightarrow 3p$ transition, still at room temperature. How can this be?]

15.4   Electric dipole radiation

So far we have treated the coefficients A_{21} and B_{21}, B_{12} as parameters which are given us. We will now see what determines them.

The easiest one to think about is B_{12}, the absorption coefficient, and the others follow from the Einstein relations.

The way an incident light beam couples to an atom is primarily through the electric field $\mathbf{E}$ of the light coupling to the electric dipole moment $\mathbf{d}$ in the atom. In an energy eigenstate, the atom has no dipole moment, but an oscillating electric dipole appears during a transition, owing to the oscillating charge distribution shown by equation (160).

Consider a single one of the electrons, at displacement $\mathbf{r}$ from the nucleus. The electron charge $-e$, together with the remaining charge $+e$ of the rest of the atom, form an electric dipole moment $\mathbf{d} = -e\mathbf{r}$. The Hamiltonian for the coupling to the electric field of the light is $H = -\mathbf{E} \cdot \mathbf{d} = e\mathbf{E} \cdot \mathbf{r}$. To find the behaviour we need to do *time dependent* perturbation theory. Without going into the full derivation, the result is that the transition rate is proportional to the *electric dipole matrix element* squared:

$$R \propto |\langle \psi_2 | e\mathbf{E} \cdot \mathbf{r} | \psi_1 \rangle|^2 \tag{174}$$

where ψ_1, ψ_2 are the atomic wavefunctions for the two states involved in the transition. The full equation can be derived using the Fermi Golden rule, it is

$$R_{1 \rightarrow 2} = \frac{\pi}{2\hbar} |\mathbf{E}_0 \cdot \mathbf{d}_{21}|^2 \delta(\hbar\omega_{21} - \hbar\omega) \tag{175}$$

where $d_{12} = \langle \psi_2 | e\mathbf{r} | \psi_1 \rangle$ is the electric dipole matrix element, and the δ-function is always integrated over in practice when we consider the spectral distribution of the incident light and the finite lifetime of the excited atomic state.

15.5   Selection Rules

For many transitions in any given atom, in fact for most of them, the electric dipole matrix element is zero! It is possible to see why without doing all the integrals. For example, if ψ_2 has the same parity as ψ_1, then since the dipole operator is an odd operator, the integral vanishes. Now, states of even l (orbital angular momentum quantum number) have even parity, and states of odd l have odd parity. [Problem: prove this. You should see immediately that you only need to consider the angular parts of the wavefunctions.] Therefore the parity change requirement leads to the *selection rule $\delta l \neq 0$* for electric dipole transitions, where l is the orbital angular momentum quantum number of the jumping electron.

Further selection rules are given in the following table, with a brief statement of the physical argument why the transition can't happen. The columns marked 'magnetic dipole' and 'electric quadrupole' are for alternative transition processes. In practice all three occur if they are allowed by the rules, but whenever electric dipole radiation is allowed its probability greatly exceeds those of the other cases. (You are only expected to know the electric dipole selection rules for finals).

	electric dipole	magnetic dipole	electric quadrupole
1.	$J = 0 \nrightarrow J = 0$	$J = 0 \nrightarrow J = 0$	$J = 0 \nrightarrow J = 0$
			$J = 1/2 \nrightarrow J = 1/2$
			$J = 0 \nrightarrow J = 1$
	and $M_J = 0 \nrightarrow M_J = 0$ when $\Delta J = 0$		
2.	$\Delta J = 0, \pm 1$	$\Delta J = 0, \pm 1$	$\Delta J = 0, \pm 1, \pm 2$
	$\Delta M_J = 0, \pm 1$	$\Delta M_J = 0, \pm 1$	$\Delta M_J = 0, \pm 1, \pm 2$
3.a	parity change	no parity change	no parity change
3.b $\implies$	$\Delta l = \pm 1$	$\Delta l = 0$	$\Delta l = 0, \pm 2$
4.	configuration change: only one electron jumps		
5.	any Δn	$\Delta n = 0$	any Δn
6.	$\Delta S = 0$	$\Delta S = 0$	$\Delta S = 0$
7.	$\Delta L = 0, \pm 1$	$\Delta L = 0$	$\Delta L = 0, \pm 1, \pm 2$

These rules have been given in order of how strict they are. The rules in the first box are strict, those in the second box are valid in the central field approximation (and therefore are fairly well obeyed for all atoms); those in the third box are only relevant for LS coupled atoms, and since LS coupling is never perfect they are not strict. 1 and 2 are about the overall angular momentum of the atom; 3 to 5 concern allowed changes in the configuration; 6 and 7 changes in the term.

Reasons for the above
1. radiation carries angular momentum ... conservation of same
2. dipole radiation carries away 1 unit of angular momentum, quadrupole 2 units of angular momentum, etc.
3. $\left\{ \begin{array}{ll} \mathrm{E}l & \text{is an even operator for even } l, \\ \mathrm{M}l & \text{is an even operator for odd } l \end{array} \right\}$, and upper limit on orbital angular momentum change is as in (2).
4. electrons are independent in central field approximation; if more than one electron changes state then the matrix element will involve a sum of terms of the form

$$|\langle \psi_f(\mathbf{r}_1) \phi_f(\mathbf{r}_2)| e\mathbf{E} \cdot \mathbf{r_1} |\psi_i(\mathbf{r}_1)\phi_i(\mathbf{r}_2)\rangle|^2 = |\langle \psi_f(\mathbf{r}_1)| e\mathbf{E} \cdot \mathbf{r_1} |\psi_i(\mathbf{r}_1)\rangle|^2 |\langle \phi_f(\mathbf{r}_2)| \phi_i(\mathbf{r}_2)\rangle|^2 = 0$$

where the second bracket is zero since ϕ_f is orthogonal to ϕ_i (when more than one of the contributions to the total product wavefunction changes).
5. there is always some radial overlap. Note, for electric dipole case, since $\Delta l \neq 0$, the l change is sufficient to cause an oscillating electric dipole so both $\Delta n = 0$ and $\Delta n \neq 0$ is ok.
6. the electromagnetic field does not couple to spin, so it can't affect the size of the spin of the atom
7. angular momentum conservation, photon angular momentum comes entirely from orbital motion of atom for electric dipole case. In magnetic dipole, by contrast, transitions are between different J levels of the same multiplet ($\Rightarrow$ microwave).

Violations
Violations occur primarily through the fact that the atom is not precisely described by the approximations. For example, '$\Delta S \neq 0$' transitions are due to the fact that LS coupling is not completely correct. Similarly, configuration mixing explains 'disallowed' changes in l. That is, a required parity change is rigorous for electric dipole radiation, but since the central field approximation

is not exact, each energy eigenstate does not correspond precisely to one configuration, but to a superposition of configurations. A transition exhibiting a disallowed change in l actually takes place by an allowed change in l in some small component in the superposition.

However, since J is always a good quantum number (in the case of no hyperfine structure), its selection rules are rigorous. With hyperfine structure, it is the selection rules on F which are most rigorous, though those on J are still very good since the nuclear magnetic moment hardly affects the electronic wavefunctions. For a $F = 0 \leftrightarrow F = 0$ transition, the atom must emit (or absorb) at least two photons, a process which has very low rate.

15.6   Do electrons jump?

It is common parlance to talk about an electron 'jumping' in an atomic transition, but there are two reasons why this only gives in incomplete picture. First, as we have seen, absorption and emission, whether driven or spontaneous, is not an abrupt process but a continuous one, whose time scale can be short (10 ns for allowed optical transition) or long (millions of years for microwave transitions). When a photon detector registers a 'click', that does not imply the atomic emission was an abrupt process, it implies that the mechanism of the detector had a short time-scale.

Also, it is not really valid to talk of 'an electron' jumping, because the electrons are identical and therefore indistinguishable particles. The phrase is a short-hand for a change of state of the *whole atom*. Indeed, if one of the single-electron wavefunctions changes, then the charge distribution in the atom changes, so all the other single-electron wavefunctions change a little as well. For example the "2s" wavefunction in the configuration $1s^2 2s 2p$ of Berylium is not the same wavefunction as the "2s" wavefunction in the configuration $1s^2 2s 3d$ of Berylium, therefore in the transition $2p \rightarrow 3d$ the whole atom is involved.

15.7   Main points

The main points of this chapter have been:
absorption or emission of light requires accelerating charge, which happens when an atom is in a superposition of energy eigenstates; Einstein A and B coefficients, their definition and the derivation of their relation; the rate equations and their use, for example in analyzing the laser; population inversion; natural linewidth and Lorentzian shape; Doppler broadening and Gaussian shape; electric dipole matrix element and selection rules.

16 Further experimental methods

These lecture notes have taken the approach mostly of describing atomic structure without reference to the experimental methods which were crucial in enabling atoms, and indeed the whole of quantum physics, to be understood in the first place, and which are crucial to permit all the understanding we now have to be put to good use in finding out about time, and stars, and quark interactions, etc.

The work-horse instruments for studying atomic spectra are the diffraction grating and the Fabry-Perot etalon. The former was discussed in chapter 3, the latter is discussed below. Both instruments, and also the Michelson interferometer, are used by astronomers to deduce the properties of stars from what is observed about their atomic constituents. For example, the speed of a star with respect to the earth, its surface temperature and magnetic field can all be deduced in the right conditions, as well of course as the atomic composition. All the instruments also feature in modern atomic physics experiments, though not always directly as spectrometers in the traditional sense. Rather, they are used to control and analyze the wavelengths of lasers, and spectroscopy is done by observing the absorption or emission when a laser beam of known and tunable wavelength illuminates an atomic sample. A high precision is needed. For example a strong (electric dipole allowed) optical transition might have an upper state lifetime or order 10 ns, and hence a linewidth of order 15 MHz. The optical region of the spectrum has frequencies of order 5×10^{14} Hz, so to locate a line to within its natural linewidth requires a relative precision of order 10^{-8}. No single instrument can achieve this, but by ingenuity and combining methods this, and even finer precision, can be achieved.

16.1 The Fabry-Perot Etalon

I will not give a full description of this instrument here, which is part of the optics course. I will highlight the main features which are relevant to its practical use.

The physical construction of a Fabry-Perot involves two main considerations: the choice of distance between the plates, and their reflectivity. These two choices directly influence the main two parameters which are relevant to the use of the etalon: its *free spectral range* and *Finesse*.

16.1.1 Free spectral range

When illuminated by monochromatic light of wavelength λ, the etalon transmits at a set of well-defined angles of incidence, and reflects at all other angles. The angles θ of high transmission are given by

$$2nd\cos\theta = p\lambda \tag{176}$$

where d is the plate separation, n is the refractive index of the material between the plates and p is an integer[16].

When performing spectroscopy, we typically have a light source emitting two or more well-defined wavelengths simultaneously. Take the example of just two wavelengths, and hence two wavenumbers $\tilde{\nu}_1 = 1/\lambda_1$, $\tilde{\nu}_2 = 1/\lambda_2$. Each will give rise to its own fringe pattern. The combined fringe pattern is easy to interpret when $\tilde{\nu}_2$ is close to $\tilde{\nu}_1$, but at certain values of $\tilde{\nu}_2 - \tilde{\nu}_1$ an ambiguity

---

[16]We are ignoring a fixed additional phase change associated with the reflections, it can be included by replacing p by $p + \epsilon$ where p is an integer and ϵ is a fixed number in the range $0 < \epsilon < 1$.

arises because the fringe patterns exactly overlap one another. This happens whenever the angles for one pattern are the same as the angles for the other, i.e.:

$$2nd\cos\theta = p_1\lambda_1 = p_2\lambda_2 \tag{177}$$

for some pair of integers p_1, p_2. The smallest separation giving such an ambiguity is when $p_2 = p_1 \pm 1$. Without loss of generality, let λ_2 be the shorter wavelength, then $p_2 = p_1 + 1$. Putting this in eq. (177) and dividing both sides by $p_1\lambda_1\lambda_2$ we have

$$\frac{1}{\lambda_2} = \frac{1}{\lambda_1} + \frac{1}{p_1\lambda_1} \tag{178}$$

$$\Rightarrow \quad \tilde{\nu}_2 - \tilde{\nu}_1 = \frac{1}{2nd\cos\theta} \tag{179}$$

The separation of wavenumbers which gives such overlapping orders for normal incidence ($\theta = 0$) is called the *free spectral range*

$$\tilde{\nu}_{\text{f.s.r.}} \equiv \frac{1}{2nd}$$

The important point is that the free spectral range is an intrinsic property of the etalon alone, it does not depend on the wavelength of the light illuminating it.

The free spectral range is both a blessing and a curse. If we have once accurately measured the f.s.r. of a given etalon, then thereafter it provides a very useful intrinsic calibration of the spectrum observed with the instrument. For example, two spectral lines from a source will give a sequence of pairs of bright fringes when observed with an etalon. It is easy to compare the observed splitting of each pair to the observed distance from one pair to the next. If the splitting is found to be a fraction α of this distance, then the wavenumber separation of the components must be $(\Delta p + \alpha)\tilde{\nu}_{\text{f.s.r.}}$, where Δp is an integer. Thus the calibration is automatically given, apart from the integer Δp. Typically we may know Δp to be zero since we have previously established that the splitting of the components is less than $\tilde{\nu}_{\text{f.s.r.}}$, by using another etalon of larger free spectral range, or by using another instrument such as a diffraction grating.

The free spectral range is a curse in the sense that it means a spectrum consisting of many components spread over more than one free spectral range will be too difficult to interpret. In other words, the Fabry-Perot etalon is a tool which is useful for
high resolution over a restricted wavelength range.
In practice an etalon is never used on its own: there has to be some other device to filter out all the light not in the small region of the spectrum we want to investigate.

To have an idea of orders of magnitude, typical etalons come with plate separations in the range 0.5 mm to 10 cm, with n in the range 1 to 1.5. Hence the most common free spectral range values are in the range

$$\tilde{\nu}_{\text{f.s.r.}} = 0.05 \text{ to } 10 \text{ cm}^{-1}, \quad c\tilde{\nu}_{\text{f.s.r.}} = 1.5 \text{ to } 300 \text{ GHz}$$

where the second set of values gives the range in frequency units. These values are tiny compared to the full range of wavenumbers or frequencies emitted by any atom, but of the same order as hyperfine structure and typical Zeeman effect observations. Hence the etalon is excellent for measuring hyperfine structure and Zeeman spectra.

16.1.2 Finesse

Finesse is a beautiful word. Experimental expertise is all about finesse. In the context of the etalon, it refers to the sharpness of the transmitted interference fringes. To be precise,

$$\text{finesse} \quad \mathcal{F} \equiv \frac{\text{Free spectral range}}{\text{FWHM of any fringe}}$$

hence large finesse is good: it implies sharp fringes and good resolution[17]. From the analysis of the optics, the following formula can be derived:

$$\mathcal{F} = \frac{\pi\sqrt{R}}{1-R} \tag{180}$$

where $R = |r|^2$ is the intensity reflectivity of the plates.

The reflectivity depends on the mirror coating, and is usually chosen as high as possible. Very high reflectivities ($R > 0.999$) can be achieved with multi-layer dielectric coatings, but there is always some absorption, which reduces the overall transmission through the etalon. The effect of the absorption will be negligible if it is small compared to $1 - R$, so we back off a little from the very highest value and $R = 0.95$ is acceptable. This gives a finesse of about 60; more generally values in the range 50 to 500 are realistic for spectroscopy purposes. However, you will be interested and I hope intrigued to know that the very highest finesse available from manufactured mirrors and coatings is in excess of 1 million.

Two wavenumbers are resolved by the etalon when the fringes they produce are separated by about a FWHM. Therefore the minimum resolvable wavenumber difference is

$$\delta\tilde{\nu}_{\min} = \frac{\tilde{\nu}_{\text{f.s.r.}}}{\mathcal{F}} \tag{181}$$

Notice that this result is consistent with the general principle given in section (3.2). Here the light bounces back too and fro between the etalon plates a number of times of order $\mathcal{F}$ (to see this, consider the loss per reflection) so the total distance available for interference is $2d\mathcal{F}$, so the general principle implies the smallest resolvable wavenumber separation is $1/2d\mathcal{F}$ in agreement with (181). [Problem: have I forgotten the refractive index? How does this influence the general principle?].

Putting together the typical free spectral range values quoted in the previous section and finesse values quoted in this section, we find that etalons can have a truly impressive resolution, down to a few MHz in frequency units.

16.2 Comparison of diffraction grating and Fabry-Perot etalon

Although we meet the concept of free spectral range in the context of the etalon primarily, the same concept applies to any spectroscopic instrument. For the diffraction grating, for example, the various orders of diffraction can produce overlapping spectra. A rough indication of the free spectral range of a grating is obtained by finding out a pair of wavelengths such that the p'th order diffraction of one overlaps the $p+1$'st order diffraction of the other:

$$d\sin\theta = p\lambda_1 = (p+1)\lambda_2$$

hence the free spectral range $\tilde{\nu}_2 - \tilde{\nu}_1$ is found to be equal to $\tilde{\nu}_1/p$, i.e. of the order of the wavenumbers involved. Hence the 'finesse' of a grating (i.e. the free spectral range divided by the smallest observable splitting) is equal to N, the number of elements in the grating, (recall (29)).

---

[17]'FWHM' is "full width at half maximum": the distance across a peak at a position half way down the peak.

A major technical limitation of both instruments is the flatness of the surfaces. For a grating we need a large area surface of flatness small compared to λ. For an etalon we need two surfaces whose area can be small but whose flatness and parallelism need to be small compared to $\lambda/\mathcal{F}$ since there are $\mathcal{F}$ successive reflections. In practice the gratings can be made sufficiently flat with dimensions up to 10 cm or so, with elements spaced by 1 micron, so having $N = 10^5$ elements. Hence the resolving power of a good grating exceeds the finesse of most etalons. The grating is used to examine the gross structure of any spectrum, and can measure separations of order 1 cm^{-1} (since grating width is of order some cm). The etalon then 'takes over' and can be used to examine each ~ 1 cm^{-1}-wide region in fine detail.

16.2.1 Choosing an etalon to do a job: an example

For example, to measure the hyperfine structure of caesium:

Problem: Measurements with a diffraction grating revealed the fine structure of the resonance lines of caesium, and a really good diffraction grating spectrometer gave a hint that there might possibly be further unresolved structure in these lines. The diffraction grating had width 5 cm and the surrounding optics were of high quality and carefully set up, so that the instrument was within a factor 2 of its theoretical limit of resolution. How can the further structure of the D1 line (wavelength 852 nm) be examined? (There are some glass Fabry-Perot etalons in the cupboard which have spacings 0.936, 1.904, 3.722 and 8.132 mm.)

Answer: The theoretical limit resolution for a 5 cm grating is of order $1/5$ cm^{-1}, so if it got within a factor 2 of this then the resolution was around 0.4 cm^{-1}. It seems therefore that the features to be resolved have splitting of this order. We had best pick an etalon whose free spectral range comfortably spans this range, and whose finesse is high, to give good resolution.

Opting therefore for a free spectral range of 2 cm^{-1}, we pick an etalon made from a glass plate with refractive index $n \simeq 1.5$, so we want thickness d given by $1/(2nd) = 2$ cm^{-1} hence $d = 0.166$ cm $\simeq 1.7$ mm. The ones we have in the cupboard have thickness 0.936, 1.904, 3.722, 8.132 mm so we pick the second one. Its free spectral range is 1.742 cm^{-1} where note that to calculate this we need to use a refractive index value good to 4 significant figures to make full use of the precision available given the finesse of 200. The index value was looked up for the wavelength in question which is 852 nm. In frequency units this free spectral range is 52.23 GHz. With a finesse of 200 we can resolve features down to 260 MHz or so.

To use the etalon, leave the diffraction grating spectrometer in place and isolate the 852 nm line, that is, only permit this line to be transmitted to the etalon by placing a slit at the grating spectrometer output (the so-called 'monochromator' arrangement). The light does not have to be collimated before passing through the etalon, but approximate collimation can serve to increase the brightness of the central few fringes at the etalon output. A lens is placed after the etalon and the fringes can be observed in the focal plane of this lens.

Further comments. The etalon revealed that there are indeed 2 components to the transition, and furthermore it looked as though both components had sub-structure but the resolution was limited by the Doppler broadening in the emission lamp of order 400 MHz. The separation of the components was found to be 9.19 ± 0.02 GHz (where note it is perfectly possible to locate the centre of a feature to within a small fraction of its width). This reveals the hyperfine structure of the ground state of caesium.

16.3 Beyond the Doppler limit: further experimental methods

The Doppler broadening at room temperature is a major limitation to the study of atomic spectra. Its influence can be minimised by any of several approaches:

1. Cooling. Since the broadening is proportional to $\sqrt{T}$, a factor 100 in T is needed to gain on order of magnitude reduction in the broadening. This means simple cooling methods are not much good. Laser cooling (see M.Phys. option) however is very successful for those atoms to which it can be conveniently applied.

2. Transverse atomic beam. Instead of a gas, we can use a beam of atoms effusing from an oven. A slit downstream serves to collimate the beam, and then we have only a small range of *transverse* velocities. We can observe the light emitted in the transverse direction, and/or illuminate the atoms with light propagating in the transverse direction, and thus avoid most of the Doppler broadening.

3. Microwaves. The Doppler shift for microwaves is in any case very small, so any method which can reveal emission or absorption of microwaves can have high precision. However, this requires further ingenuity because at room temperature atoms on average neither emit nor absorb microwaves, simply because the thermal populations of the levels mean that any pair of states with a splitting small compared to $k_B T$ will be very close to equally populated.

4. Saturation spectroscopy and other 'Doppler-free' methods. These involve some clever tricks using more than one laser beam to select out atoms from a vapour cell having a narrow range of velocities, or to cancel the Doppler shift in a two-photon transition: see M.Phys option!

Of these, 1 and 2 are simple enough so you should be aware of them. The only one of the others which I will comment further on here is 3.

Whenever the Doppler broadening is avoided, the experimental resolution must be limited by other considerations. There is always the natural linewidth of the atomic transitions, and the resolution of the instruments being employed, but sometimes both of these are very small (e.g. microwave experiments). In this case the limit on resolution of the experiment is often set by the *transit time*. This is the time spent by the atom in the radiation field which excites it. By straightforward Fourier analysis, if an atom only experiences an oscillating field for a time τ, then the frequency at the atom is only well-defined to a precision of order $1/\tau$, so spectral features can only be resolved to this precision.

Microwave spectroscopy of atoms has been carried out to great success by the atomic beam method. The apparatus consists of a pair of back-to-back Stern Gerlach apparati, with a region of uniform magnetic field in between, where microwaves can be applied. The diagram was given in the lecture, see also Haken and Wolf and Corney. A microwave transition is detected if it changes the direction of the magnetic dipole moment of the atom, because then the atom takes a different path through the 2nd Gerlach apparatus. The change in path is detectable since it determines whether or not the atom hits a small detector (hot wire filament). The resolution is limited by transit-time broadening: a typical value is given by a microwave region 10 cm long with atoms in the beam at thermal velocities of order 300 metres per second, hence $\tau \simeq 0.3$ ms so the spectral features have width 3 kHz.

A modern atomic clock achieves transit time of 1 s. This is not possible with room temperature atoms: the detection region would have to be too long. It is achieved by throwing a sample of ultra-cold atoms vertically upwards in a vacuum chamber about 1 metre high. The cold atoms pass twice through a small microwave cavity, the two passes separated in time by about 1 s. The observed spectral features have a broad envelope of width 10 Hz, with superimposed on it a set of

fringes in frequency-space separated by 1 Hz: this is a "Young's two-slit" arrangement in frequency space, it enables the transition frequency to be pin-pointed very accurately. (Here is yet another Nobel prize—in 1989 to Norman Ramsey for his invention and high-precision experimental use of this method.)

To get even higher precision, the atoms must be held in a trap. The best precision is now available from ion traps which tightly confine charged atoms for days or weeks at a time. The precision in this case is limited by technical considerations other than transit time broadening.