PHY332

Atomic and Laser Physics

A.M. FOX

Autumn Semester

2016

Course synopsis

Outline syllabus

PART I: ATOMIC PHYSICS. Quantum mechanics of the hydrogen atom. Radiative emission by atoms and selection rules. Shell model and alkali spectra. Angular momentum coupling. Helium and the exchange energy. Spin-orbit coupling and spectral fine structure. The Zeeman and Stark effects.

Part II: LASER PHYSICS. Stimulated emission. Einstein's A and B coefficients. Population inversion. Laser modes. Examples of lasers systems and their applications. Laser cooling of atoms.

Lecture Notes

1. Introduction and revision of hydrogen
2. Radiative transitions
3. The shell model and alkali spectra
4. Angular momentum
5. Helium and exchange symmetry
6. Fine structure
7. External fields: the Zeeman and Stark effects
8. Lasers I: stimulated emission
9. Lasers II: cavities and examples
10. Laser cooling of atoms

Assessment

The course is assessed by **Homework (15%)** and **Exam (85%)**. Students frequently ask whether all the material in these lecture notes is examinable. The answer to this is no, but not in simple way. Some derivations are clearly included for pedagogical purposes, and you will not be asked to reproduce them in the exam. Most of these are given in the appendices, which are not included in the printed notes, and can be downloaded from the web. In other cases, I might expect you to be aware of the consequences of a detailed derivation, although I would not expect you to reproduce the derivation in the exam. One example is the derivation of the spin-orbit perturbation in Chapter 7. Here, I expect you to know that the perturbation is proportional to $\boldsymbol{l} \cdot \boldsymbol{s}$, and that it increases with Z, but I have never asked for a detailed derivation in the exam. Similarly, I have never asked for the evaluation of the Stark shifts by perturbation theory considered in Appendix D in the exam, but I do expect you to know why the quadratic Stark shift varies in magnitude from transition to transition, and why some transitions show a linear Stark shift. I will try to make these distinctions plain as I go through the lectures. Therefore, if you want to save yourself work at revision time, come to the lectures!

Online resources

Most of the information in these notes is available on the course web page. The web address of the course page is http://www.mark-fox.staff.shef.ac.uk/PHY332/. The appendices are not included in the printed notes. They can be downloaded from the www page.

Recommended books

- Bransden, B.H. and Joachain, C.J., *Physics of Atoms and Molecules*, (2nd edn, Prentice Hall, 2003)
- Demtröder, W., *Atoms, Molecules and Photons*, (Springer-Verlag, 2006)
- Haken, H. and Wolf, H.C., *The Physics of Atoms and Quanta*, (7th edn, Springer-Verlag, 2005)
- Hertel, I.V. and Schulz, C.-P., *Atoms, Molecules and Optical Physics Vol. 1: Atoms and Spectroscopy*, (Springer-Verlag 2015)
- Hooker, S. and Webb, C., *Laser Physics* (Oxford, 2010): introductory course sections (see p. vi)

Also useful

- Eisberg and Resnick, *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles* (Wiley, 1985)
- Hecht, *Optics* (3rd edn, Addison Wesley, 1998), Section 13.1
- Phillips, *Introduction to Quantum Mechanics* (Wiley, 2003)
- Smith and King, *Optics and Photonics* (Wiley 2000), chapters 15–17
- Wilson and Hawkes, *Optoelectronics, an introduction*, (3rd edn, Prentice Hall (1998): Chapters 5–6 on laser physics

More advanced texts

- Foot, *Atomic Physics* (Oxford, 2005)
- Silfvast, *Laser Fundamentals* (2nd edition, Cambridge, 2004)
- Svelto, *Principles of Lasers* (4th edn, Plenum, 1998)
- Woodgate, *Elementary Atomic Structure* (Oxford, 1980)
- Yariv, *Optical Electronics in Modern Communications* (5th edition, Oxford, 1997)

Acknowledgements

These notes are available publicly on the www, and I am very grateful to receive comments from colleagues around the world on their content. I would like to thank Dr André Xuereb from the University of Malta for his comments on the 2013 version of the notes.

Contents

Chapter 1

Preliminary concepts

Atomic Physics is the subject that studies the inner workings of the atom. It remains one of the most important testing grounds for quantum theory, and is therefore a very area of active research, both for its contribution to fundamental physics and to technology. Furthermore, many other branches of science rely heavily on atomic physics, especially astrophysics, laser physics, solid-state physics, quantum information science, and chemistry. So much so, that Richard Feynman once wrote:[1]

> If, in some cataclysm, all scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the *atomic hypothesis* (or atomic fact, or whatever you wish to call it) that *all things are made of atoms - little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another.* In that one sentence you will see an *enormous* amount of information about the world, if just a little imagination and thinking are applied.

Atoms consists of a nucleus surrounded by electrons in quantized orbits. The main task of the subject of atomic physics is to determine the energies and wave functions of the quantised electron states, setting aside the study of the nucleus to the subject of nuclear physics. Before we set about this task, it is first necessary to cover a number of important basic concepts and definitions. These are covered in this present chapter.

1.1 Quantised energy states in atoms

The first basic concept we need is that of **bound states**. Atoms are held together by the attractive force between the positively-charged nucleus and the negatively-charged electrons: the electrons are *bound* to the atom, rather than being *free* to move though space. In the limit where the electron is very far away from the nucleus, the attractive force is negligible; the electron is free to move without any influence from the nucleus with velocity $\boldsymbol{v}$, as illustrated schematically in Fig. 1.1(a). It is natural to define the energy E of this free (or *unbound*) state as being zero when $\boldsymbol{v} = 0$. When the electron moves closer to the nucleus, it begins to experience the attractive force, leading to the formation of a stable bound state as illustrated in Fig. 1.1(b). The energy of the bound state is lower than that of the free electron, since it requires energy to pull the electron away from the nucleus. The amount of energy required is called the **binding energy** of the electron. With our definition of $E = 0$ corresponding to the unbound state, the absolute energy E of the bound state must be negative, with the binding energy equal to $-E = |E|$.

The early understanding of the atom was built around the analogy with the solar system, where the planets orbit around the sun under the influence of the attractive gravitational force. While it will not be appropriate to push this analogy too far on account of the need to use quantum mechanics rather than Newtonian mechanics to describe the motion, it does provide a useful starting point. In the same way that the planets arrange themselves into orbits at varying radii from the sun, the electrons in an atom are arranged in a series of quantised states around the nucleus. The planets nearest the sun are very strongly bound and have small radii with fast periods. The outer planets, by contrast, are less strongly bound, and have large radii and long periods. Similarly, the electrons are arranged into orbital **shells** around the nucleus. The electrons nearest the nucleus are very strongly bound, while those further away are more

[1] R.P. Feynman, *The Feynman Lectures on Physics*, Vol. I, §1-2, Addison-Wesley 1964

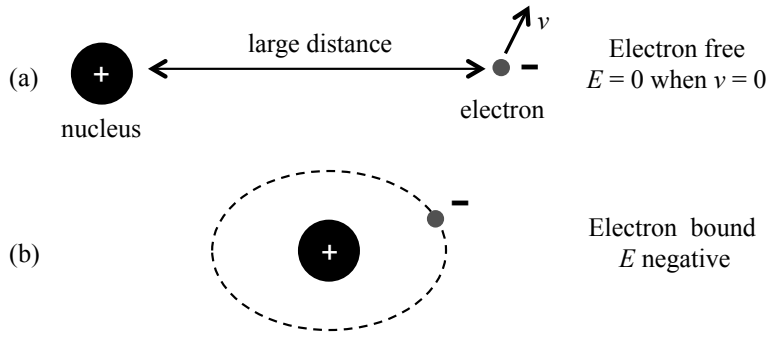

Figure 1.1: (a) Unbound state with the electron far from the nucleus. The electron moves freely with velocity $\boldsymbol{v}$, independently of the presence of the nucleus. (b) Bound electron state with negative energy.

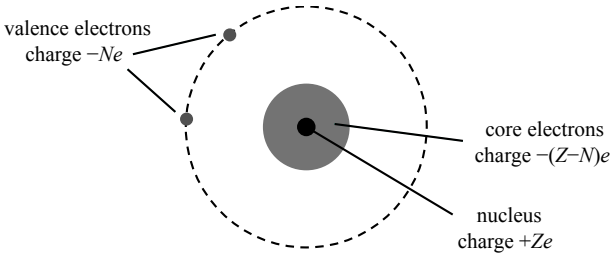

Figure 1.2: Arrangement of electrons into core and valence shells within a neutral atom of atomic number Z with N valence electrons.

weakly bound. The arrangement of the electrons within these quantised shells around the nucleus is the basis of the **shell model** of the atom discussed in Chapter 4.

Elements are identified by their atomic number Z, which defines the number of protons in the nucleus. The nucleus also contains neutrons, but these are uncharged and only affect its mass.[2] Since the charges of protons and electrons are identical, the charge of the nucleus is equal to $+Ze$, where e is the electron charge. Free atoms are normally found in a neutral electrical state, which means that they have Z bound electrons. (Charged atoms are discussed in Section 1.2 below.) The electrons in the innermost quantised states are very strongly bound and are called **core electrons**. The remaining electrons in the outermost shell are called **valence electrons**. It is these valence electrons that take part in chemical bonding, with the number N of valence electrons determining the chemical valency of the atom. Assuming that $Z > N$, which is the case for all neutral atoms apart from hydrogen and helium, the remaining $(Z - N)$ electrons are in inner, core shells, as illustrated in Fig. 1.2. These core electrons are very strongly bound and can only be accessed by using high energy X-ray photons, as discussed in Section 4.4. The optical spectra of the atom are determined by the valence electrons, which are therefore the main focus of the subject of atomic physics.

The energies of bound states in atoms are frequently quoted in **electron volt** (eV) units. 1 eV is the energy acquired by an electron when it is accelerated by a voltage of 1 Volt. Thus 1 eV $= e$ J, where $-e = -1.6 \times 10^{-19}$ C is the charge of the electron. This is a convenient unit, because the binding energies of the valence electrons in atoms are typically a few eV. The core electrons, however, have much larger binding energies, typically in the keV range for atoms with large Z.

1.2 Ionization states, and spectroscopic notation

In the previous section we considered the case of a neutral atom in which Z electrons are bound to a nucleus containing Z protons. Charged atoms also exist in which the number of electrons is different to Z. Such charged atoms are called **ions**. In atomic physics we deal almost exclusively with positively charged ions, in which the number of bound electrons is less than Z. In chemistry, however, it is also necessary to consider negative ions in which the atom binds more than Z electrons.

[2] Neutrons also contribute to the spin of the nucleus. There are two known nuclei that do not contain any neutrons. The most obvious example is ^{1}H. The other is ^{2}He, which is highly unstable.

Atom / ion	Spectroscopic notation	Number of electrons
Na	Na I	11
Na^+	Na II	10
Na^{2+}	Na III	9
⋮	⋮	
Na^{10+}	Na XI	1

Table 1.1: Ionization states of the element sodium (chemical symbol Na), which has an atomic number of 11.

The **ionization energy** of an atom (also sometimes called the **ionisation potential**) is defined as the lowest energy required to remove an electron. The electrons are bound to the atom in shells with different quantised binding energies, and the ionization energy is equal to the binding energy of the least strongly bound electron. In practice, this will be one of the valence electrons.

Hydrogen is the first element and has $Z = 1$. Since it only binds one electron, it only has one ionization energy. All other atoms have more than one bound electron, and therefore have more than one ionization energy. An atom with atomic number Z has Z **ionization states**, and hence Z ionization energies. The nth ionization energy is defined as the energy required to remove the nth electron from the atom, according to the following sequence:

$$\begin{aligned} \mathrm{A} &\rightarrow \mathrm{A}^+ + \mathrm{e}^- && \text{1st ionization energy} \\ \mathrm{A}^+ &\rightarrow \mathrm{A}^{2+} + \mathrm{e}^- && \text{2nd ionization energy} \\ &\vdots && \vdots \\ \mathrm{A}^{(Z-1)+} &\rightarrow \mathrm{A}^{Z+} + \mathrm{e}^- && Z\text{th ionization energy} \end{aligned}$$

where A^{n+} represents an atom, A, that has lost n electrons from the neutral state, with A^{Z+} corresponding to an isolated nucleus. Each ionization state has a unique spectrum, which allows the atom to be identified from analysis of its spectral lines.

In normal laboratory conditions at temperature T, with $T \sim 300\,\mathrm{K}$, the thermal energy $k_\mathrm{B}T$ is significantly smaller than the first ionization energy of the atom. This means that atoms are normally in the neutral state. In order to study ions, we either have to raise the temperature significantly (e.g. in a flame), or we have to deliberately strip off the electrons (e.g. in a collision with another charged particle in a discharge tube). In astrophysics, however, we study the spectra of atoms in stars, where the temperature is always very high, and highly ionised states are routinely found.

Astronomers have been studying the spectra of atoms and ions for a long time, using the characteristic spectral lines of the elements to determine the composition of stars. Joseph von Fraunhofer (1787–1826), among others, carried out a systematic study of the Sun's spectrum in the 19th century, and identified a number of absorption lines with those of certain elements in the periodic table. In order to categorize the spectral lines, **spectroscopic notation** was introduced to identify the different ionization states of the atoms. In this notation, the Nth ionization state of atom A is written A (N+1), where (N+1) is written in capital Roman numerals. Thus A I is the neutral state of the atom, A II is the first ionization state A^+, and so on. This spectroscopic notation is widely used in astrophysics, and is also used in important databases of atomic physics. (See Section 1.5.) Table 1.1 shows how the notation is applied to the element sodium (chemical symbol Na), which has an atomic number of 11.

1.3 Ground states and excited states

An atom with atomic number Z is made by starting with the isolated positive nucleus and then adding Z electrons until electrical neutrality is achieved. As mentioned in Section 1.1 and discussed in detail in Chapter 4, the quantised electron states are arranged in shells around the nucleus. The Pauli exclusion principle, which will be discussed in Chapters 4 and 6, dictates that each shell can only hold a strictly limited number of electrons. The electrons therefore fill up the shells one by one, in sequence of increasing energy, moving to the next one after a particular shell is full. Eventually all the electrons have been bound.

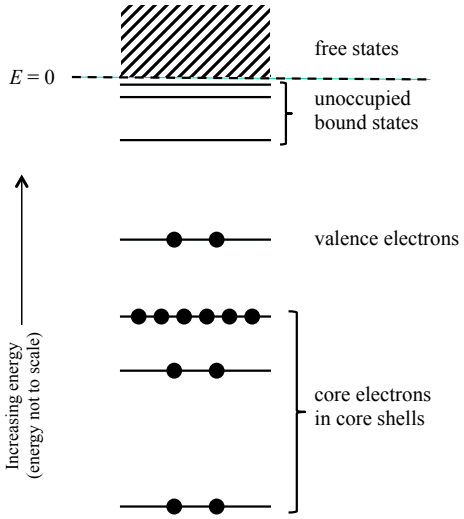

Figure 1.3: Arrangement of the energy levels and electrons in the ground state of an atom. The electrons fill up the atomic shells in order of increasing energy until all the electrons have been accounted. The shading for the free states indicates that the energy is not quantised: it forms a continuum. The diagram is drawn for the case of the neutral Mg atom ($Z = 12$), which has 12 electrons. Note that the energy scale is not linear. The core shells are very strongly bound, and their large negative energies would be way off the page on a linear scale. These core electron states are usually omitted from atomic energy level diagrams.

The final state of the atom with its electrons filling up the lowest available energy shells is called the **ground state** of the atom.

The ground state of a typical atom is shown schematically in Fig. 1.3. As before, we assume that there are N valence electrons, and therefore $(Z - N)$ core electrons. The diagram is drawn for the specific case of the neutral magnesium atom, where $Z = 12$ and $N = 2$. Each horizontal line indicates a quantised energy state, and the vertical axis is energy. The zero of energy is defined as the point at which the electron is free, with all the quantised bound states having negative energy, as discussed in Section 1.1. The shading for the free states indicates that the energy is not quantised. The electron is free to move with arbitrary kinetic energy, and so can have any positive energy. The free states are therefore said to form a **continuum**: there is a continuous spectrum of energies that are possible, with no breaks due to quantization.

It is important to note that energy axis in Fig. 1.3 is not linear. The core shell states have very large negative energies, and should really be way off the bottom of the page. Since the core electrons play no part in the optical spectra, they are usually omitted from atomic energy level diagrams, and this will be the policy adopted from here onwards, unless we are specifically considering the core electrons, as in Section 4.4.

There is an infinite number of quantised bound states in an atom, but only a small number (the ones with lowest energy) are occupied in the ground state configuration of the atom. All of the other states lie at higher energy. The **excited states** of the atom are obtained by promoting valence electrons to these unoccupied states at higher energy. If there is more than one valence electron, then the excited states are obtained by promoting just *one* of the valence electrons to a higher energy state, as shown in Fig. 1.4. There are an infinite number of these excited states, although we usually only need to consider the first few to explain the most important features of the optical spectra. The large number of other excited states at higher energies are increasingly more weakly bound, and eventually merge into the continuum of free states available to unbound electrons. This means that the infinitieth excited state corresponds to the ionization limit, which provides a method to define the energy of the ground state electron configuration. This energy is identified in Fig. 1.4, and can be determined experimentally by measuring the first ionization energy of the atom.

The energy gap between the ground state of an atom and its first excited state is typically much larger than the thermal energy $k_{\mathrm{B}}T$ at room temperature. This means that the atom will normally be in its ground state. In order to promote the atom to its excited states, energy must be imparted to it. This is typically done by placing the atom in discharge tube, and exciting it by collisions with electrons flowing down the tube when a voltage is applied across it. The atom can also be promoted to a specific excited

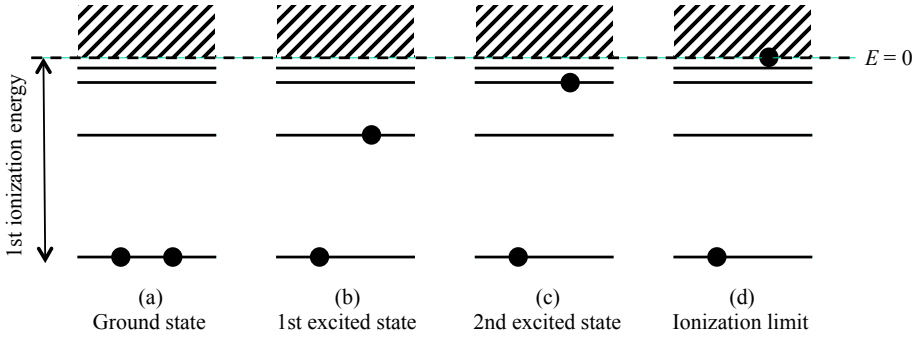

Figure 1.4: Ground and excited states of an atom with two valence electrons. (a) Ground state. (b) First excited state. (c) Second excited state. (d) Ionization limit, equivalent to the infinitieth excited state. Note that the ground state is the same as in Fig. 1.3, except that the core shells are no longer shown.

state by absorption of a photon. (See Section 1.4 below.)

For atoms that have two or more valence electrons, it is a reasonable question to ask why we only consider excited states in which only one electron is promoted to higher energy. For example, in Fig. 1.4 the 2nd excited state is the one with one electron in the lowest level, and the other in the third, rather than the state with both electrons in the second level. The answer is that it costs more energy to promote both electrons than to completely remove the first electron: the ionised state therefore has a lower energy than the unionised one with two electrons in higher levels. It is therefore easier to ionise the atom that to excite both electrons simultaneously.

The state of the atom after one electron has been removed corresponds to the singly-charged ion A^+. The method of defining a ground state and excited states starts again for this ion, with the ground state of the ion corresponding to the ionization state of the neutral atom. For example, the ionization limit of the neutral helium atom ($Z = 2$) corresponds to the ground state of the He^+ ion. (See discussion of Fig. 6.2 in Chapter 6.) If the atom has more than two electrons, this process keeps repeating itself, with the ground state of the ion A^{n+} corresponding to the ionization limit of the ion $A^{(n-1)+}$. Each ionization state has its own characteristic sequence of energy levels, which can be determined by analysis of the optical spectra. As mentioned above, this means that the ionization states of the different elements have unique spectra, which enables them to be identified, especially in astrophysics. The wavelengths of the spectral lines of the ionization states of the elements are catalogued in the NIST atomic physics database. (See Section 1.5.)

The correspondence between the ionization limit of one ionization state and the ground state of the next one is shown in Fig. 1.5. It is apparent from this diagram that the definition of $E = 0$ is a *relative* one: $E = 0$ for one ionization state corresponds to a negative energy for the next one. (This distinction does not apply, of course, to hydrogen, as it only has one electron.) In absolute terms, the true zero of energy should be defined as the state with all Z electrons stripped from the nucleus. For a multi-electron atom, this would mean that the ground state of the neutral atom, together with its excited states, all have large negative energies in absolute terms. However, since the energies of the core electrons remain constant while the valence electrons are excited, it makes sense to subtract them, defining the zero of energy for each ionization state as the energy to remove the first valence electron.

1.4 Atomic spectroscopy

We can gain a great deal knowledge about atoms from studying the way light interacts with matter, and in particular from measuring atomic spectra. Optics has therefore played a key role in the development of atomic physics. The extreme precision with which optical spectral lines can be measured makes atomic physics the most precise branch of physics. For example, the frequencies of the spectral lines of hydrogen have been measured with extremely high accuracy, permitting the testing of small but important quantum phenomena that are normally unobservable.

The basis for atomic spectroscopy is the measurement of the energy of the photon absorbed or emitted when an electron jumps between two quantized bound states, as shown in Fig. 1.6. These are called optical **transitions**. The frequency (ν) of the photon (and hence its wavelength, λ) is determined by the

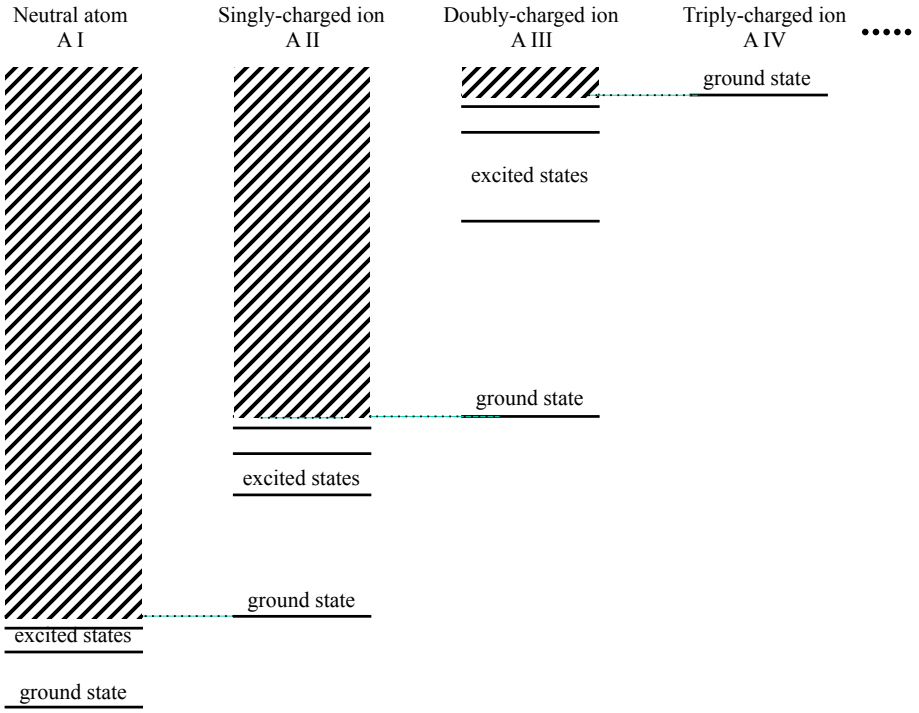

Figure 1.5: Correspondence between the ionization limit of an atom or ion, and the ground state of the next ion in the sequence. Spectroscopic notation is used for the different ionization states.

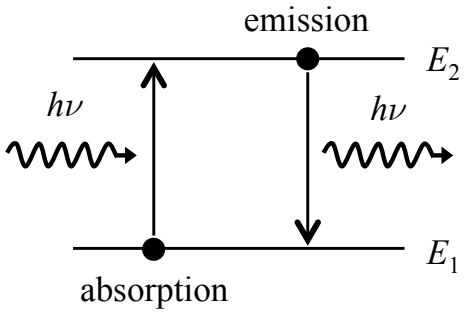

Figure 1.6: (a) Absorption and emission transitions between two quantised energy states. (b) Comparison of absorption and emission spectra of an atom.

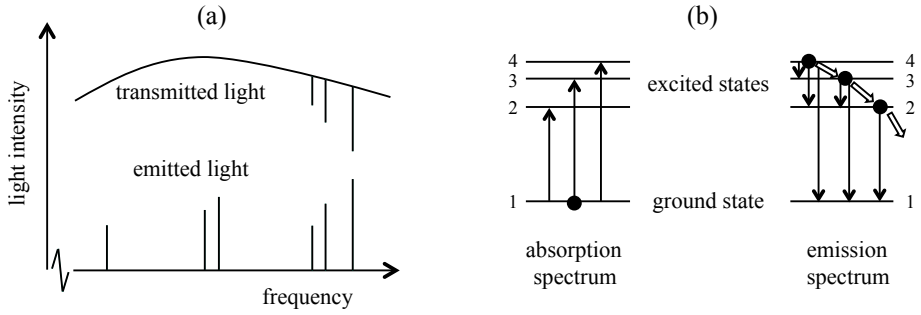

Figure 1.7: (a) Absorption and emission line spectra. For the absorption spectrum, a light source with a continuous spectrum in shone on the atoms, and the transmitted intensity is recorded as a function of frequency. The dips correspond to absorption transitions. (b) Comparison of absorption and emission transitions. The open arrows starting from the third excited state (level 4) indicate an electron cascade in which three photons of different frequencies are emitted.

difference in energy of the two levels according to:

$$h\nu = \frac{hc}{\lambda} = E_2 - E_1 \,, \tag{1.1}$$

where E_1 and E_2 are the energies of the lower and upper levels respectively, and h is Planck's constant. If the electron is initially in the lower level, it can only be promoted to the higher level if there is radiation incident on the atom. The radiation must contain photons with frequency given by eqn 1.1, and conservation of energy requires that one of these photons is removed from the beam as the electron makes its jump upwards. This is the process of **absorption**. By contrast, if the electron is initially in the upper level, then it can spontaneously drop to the lower level without the need of an external radiation field. A photon with frequency given by eqn 1.1 is emitted while the electron drops down, and the process is therefore called **emission**.[3]

The bound states of atoms have quantized energies, and so the absorption and emission frequencies that are observed from a particular atom are discrete. The **absorption spectrum** can be measured by illuminating the atoms with a continuous range of frequencies, and analysing the intensity that gets transmitted. Dips in the transmitted intensity will be observed at the frequencies that satisfy eqn 1.1, as shown schematically in Fig. 1.7(a). The **emission spectrum** can be observed by measuring the light emitted from a gas of atoms that has been excited in a discharge tube. (See Fig. 1.8.) Peaks will occur at the frequencies that obey eqn 1.1, as indicated in Fig. 1.7(a). The factors that determine the width of these peaks will be discussed in Section 3.7. At this stage, all we need to know is that the width is usually very much smaller than the centre frequency (e.g. width $\sim 10^9$ Hz as opposed to a centre frequency of $\sim 10^{14}$ Hz.) The emission peaks therefore usually just look like vertical lines unless a very high resolution spectrometer is used. The same applies to the absorption dips, which just look like vertical downward lines in the transmitted intensity. For this reason, we refer to the absorption and emission spectra of atoms as **line spectra**, with the individual transitions called absorption or emission **lines**, as appropriate. This contrasts with the spectra of solids or molecules, which can contain absorption or emission **bands**, in which a continuous range of frequencies is absorbed or emitted.

In Section 1.3 we have seen that an atom is normally found in its ground state. This means that the absorption and emission spectra of a specific ionization state of an atom show some differences. In particular, the emission spectrum is richer. This can be understood by reference to Fig. 1.7(b). In the absorption spectrum, we can only observe the transitions that start from the ground state, that is from level 1 (the ground state) to levels 2, 3, and 4. In the emission spectrum, the same transition frequencies will be observed, but there will also be others, corresponding to transitions between excited states. These transitions are possible because the electron in an excited state can drop to any level that is at lower energy, not just the ground state.[4] Photons at the intermediate frequencies are then emitted as the electron cascades down to the ground state. Thus, for example, an electron starting in level 4 can drop

[3] The process should, more accurately, be called *spontaneous* emission to distinguish it from *stimulated* emission. This distinction will be clarified in Chapter 9.

[4] This statement will be qualified in Chapter 3 where we shall see that not all transitions are possible: only those that satisfy the selection rules. These selection rules affect both absorption and emission spectra equally, and so the general point being made here about the richer nature of the emission spectrum is still valid.

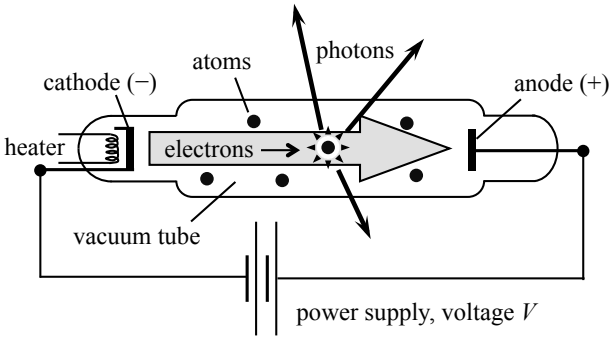

Figure 1.8: Electrical discharge tube for observing atomic emission spectra.

to any of levels 1, 2 or 3, and an electron starting in level 3 can drop to either of level 1 or 2. Therefore, for the case shown in Fig. 1.7(b), there would be three lines in the absorption spectrum, but six in the emission spectrum. The six emission lines would include the three observed in absorption, plus three others, as shown in Fig.1.7(a).

Spectroscopists measure the wavelength of the photon emitted in an optical transition, and hence deduce energy differences. The absolute energies of the quantised bound states are determined by fixing the energy of one of the levels by additional methods, and then determining the energies of the others relative to it. The simplest strategy is usually to determine the energy of the ground state (e.g. by measuring the ionization energy), and then to use it as a reference for the excited states, deducing their energies from the appropriate spectral lines. There will, of course, be many lines in the spectrum, and the individual transitions have to be identified by a process of logical deduction. For example, in Fig. 1.7 it is obvious that the three lines with highest frequency in the emission spectrum terminate on the ground state. This is confirmed by the fact that they also appear in the absorption spectrum. The next two lines terminate on the first excited state, and the sixth line terminates on the second. From this information, it is possible to work out the energies of the first, second, and third excited states relative to the ground state. As discussed in Section 1.3, the energy of the ground state relative to the ionization limit is the natural reference point for the atom.

The larger number of lines in the emission spectrum makes it more interesting to investigate. Moreover, it is usually easier to measure emission than absorption in the laboratory, as all that is needed is a **discharge tube**. In such a device, a vacuum tube with electrodes at both ends is filled with a gas of the atoms under study, as shown in Fig. 1.8. The negative electrode (the cathode) is heated to eject electrons, which then flow as a current to the positive electrode (the anode) when an external voltage V is applied. The atoms are excited by collisions with the electrons and emit photons as they relax to the ground state, either directly, or in a cascade. The maximum energy that can be imparted to the atom is equal to eV, and this determines the states that can be accessed. If eV is larger than the ionization energy, ions will be present in the tube, and their chracteristic spectra will also be observed.

The fact that each atom has a unique set of quantised energy levels, both in its neutral and ionised states, means that every element has a unique set of spectral lines, thereby providing a method to identify elements from their characteristic spectra. This technique is used extensively in astrophysics, and the discovery of the element helium is an interesting example. The Sun behaves as a source of black-body radiation with a characteristic temperature of around 6,000 K. Detailed analysis of the spectrum revealed absorption lines originating from the atoms in the Sun's atmosphere. Most of the absorption lines could be matched up to known spectra, for example, to hydrogen, which is the most abundant element in the Sun. However, the line at 587.49 nm could not be explained, and so was attributed to a new, unknown element. The element was named "helium", after *helios* in Greek, meaning the Sun. It was only several years later that helium was isolated on Earth, and the mystery line at 587.49 nm identified. We now know that helium is element number 2, and is present in large quantities in the Sun as the product of hydrogen fusion. Similar techniques are routinely used to determine the composition of other stars. Moreover, the fact that the spectra of atoms are identical throughout the universe provides a method for measuring the velocity of stars and galaxies from the Doppler shift of hydrogen lines.[5]

[5]This method relies on noticing that the special lines are all shifted in the same way, as predicted by Doppler theory. The fact that the rest-frame frequencies of hydrogen are identical in distant galaxies to those measured on Earth is one of the best tests that the laws of physics are the same throughout the whole universe.

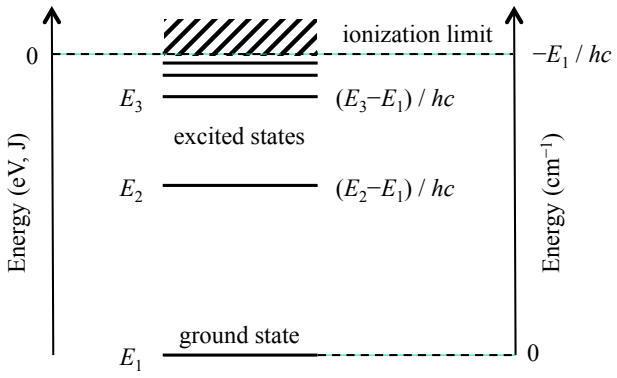

Figure 1.9: Different conventions for specifying atomic energies. On the left, we define $E = 0$ by the ionization limit, so that all the bound-state energies E_n are negative. On the right, we define $E = 0$ by the ground state, so that all the excited state energies E_n for $n > 1$ are positive. The convention on the right is the one frequently used in atomic databases, with the excited-state energies specified in cm^{-1}.

1.5 Spectroscopic energy units, and atomic databases

The close connection between atomic line spectra and the underlying level structure of the atom makes it convenient to use wave-number units (cm^{-1}) to specify the energies of the quantized bound states. The **wave number** $\overline{\nu}$ is the reciprocal of the wavelength of a photon with energy E. It is defined as follows:

$$\overline{\nu} = \frac{1}{\lambda} = \frac{\nu}{c} = \frac{E}{hc} \,. \tag{1.2}$$

The S.I. unit for wave number is m^{-1}. However, atomic spectroscopists usually use cm^{-1}, in which case it is necessary to specify λ in cm and c in cm/s in eqn 1.2. Note that $1\,\mathrm{cm}^{-1} = 100\,\mathrm{m}^{-1}$: the cm^{-1} is a *larger* unit by a factor of 100. The conversion factor to the other convenient unit for atomic levels, namely the electron Volt, is:

$$1\,\mathrm{eV} = (e/hc)\,\mathrm{cm}^{-1} = 8066\,\mathrm{cm}^{-1} \,. \tag{1.3}$$

Note again that it is necessary to use c in cm/s here (i.e. $c = 2.998 \times 10^{10}\,\mathrm{cm/s}$) to get the conversion to cm^{-1} correct.

Wave-number units are particularly convenient for atomic spectroscopy. This is because they dispense with the need to introduce fundamental constants in our calculation of the wavelength. Thus the wavelength of the radiation emitted in a transition between two levels is simply given by:

$$\frac{1}{\lambda} = \overline{\nu}_2 - \overline{\nu}_1 \,, \tag{1.4}$$

where $\overline{\nu}_2$ and $\overline{\nu}_1$ are the energies of upper and lower levels, respectively, in cm^{-1} units, and λ is measured in cm.

The convenience of wave-number units means that most professional databases use them to specify atomic energies. Moreover, these databases also usually use the ground-state level as the reference point, rather than the ionization limit. This point is clarified in Fig. 1.9, where the two different definitions of energies are compared. On the left, we have the convention that has been followed so far, following Section 1.1, where $E = 0$ is defined as the ionization limit, and all the bound-state energies E_n are negative. On the right, we have the alternative system used by spectroscopists, where $E = 0$ corresponds to the ground-state level. In this convention, the excited-state energies are positive, and specified in wave-number units relative to the ground-state. The ionization limit in cm^{-1} is then $-E_1/hc = +|E_1|/hc$.

The National Institute of Standards and Technology (NIST) in the United States maintains a particularly important on-line resource of atomic data. An extremely detailed database is provided for the use of professional research scientists, together with a simpler resource entitled the *Handbook of Basic Atomic Spectroscopy Data*.[6] Both databases use the system on the right of Fig. 1.9, with the default unit being cm^{-1}. The *Handbook* includes data for the neutral and singly-charged ion, while the professional database includes all the known ionization states. The ionization states are specified in the spectroscopic notation introduced in Section 1.2. (See, for example, Table 1.1.)

[6]The professional database may be found at http://www.nist.gov/pml/data/asd.cfm, while the *Handbook* may be found at http://www.nist.gov/pml/data/handbook/index.cfm.

	Energy scale eV	Energy scale cm^{-1}	Contributing effects
Gross structure	1 – 10	10^4 – 10^5	electron–nuclear attraction electron-electron repulsion electron kinetic energy
Fine structure	0.001 – 0.01	10 – 100	spin-orbit interaction relativistic corrections
Hyperfine structure	10^{-6} – 10^{-5}	0.01 – 0.1	nuclear interactions

Table 1.2: Rough energy scales for the different interactions that occur within atoms. The numerical values apply to the valence electrons.

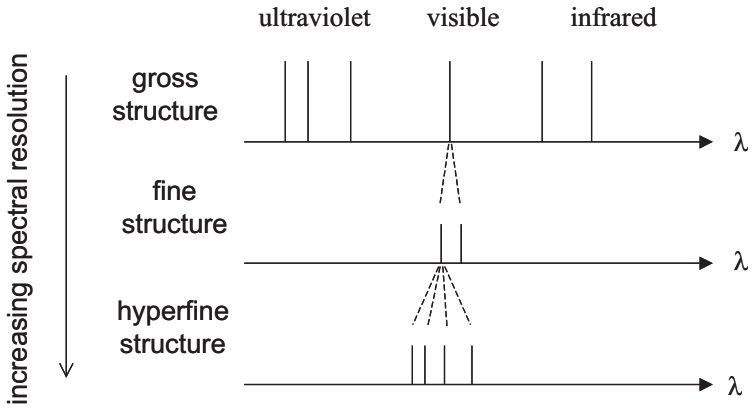

Figure 1.10: Hierarchy of spectral lines observed with increasing spectral resolution.

1.6 Energy scales in atoms

In atomic physics it is traditional to order the interactions that occur inside the atom into a three-level hierarchy according to the scheme summarized in Table 1.2. The effect of this hierarchy on the observed atomic spectra is illustrated schematically in Fig. 1.10.

Gross structure

The first level of the hierarchy is called the **gross structure**, and covers the largest interactions within the atom, namely:

- the kinetic energy of the electrons in their orbits around the nucleus;
- the attractive electrostatic potential between the positive nucleus and the negative electrons;
- the repulsive electrostatic interaction between the different electrons in a multi-electron atom.

The size of these interactions gives rise to energies in the 1–10 eV range and upwards. They thus determine whether the photon that is emitted is in the infrared, visible, ultraviolet or X-ray spectral regions, and more specifically, whether it is violet, blue, green, yellow, orange or red for the case of a visible transition.

Fine structure

Close inspection of the spectral lines of atoms reveals that they often come as multiplets. For example, the strong yellow line of sodium that is used in street lamps is actually a doublet: there are two lines with

wavelengths of 589.0 nm and 589.6 nm. This tells us that there are smaller interactions going on inside the atom in addition to the gross-structure effects. The gross-structure interactions determine that the emission line is yellow, but fine-structure effects cause the splitting into the doublet. In the case of the sodium yellow line, the fine-structure energy splitting is 2.1×10^{-3} eV or 17 cm^{-1}, which is smaller than the average transition energy (2.104 eV) by a factor of $\sim 10^{-3}$.

The main cause of fine structure is interactions between the spin of the electron and its orbital motion. Electrons in orbit around the nucleus experience a magnetic field which is internal to the atom, and this then interacts with the spin of the electron, as will be explained in Chapter 7. The **spin-orbit** energy can be deduced by measuring the fine structure in the spectra, and in this way we can learn about the way the spin and the orbital motion of the atom couple together. In more advanced theories of the atom (e.g. the Dirac theory), it becomes apparent that the spin-orbit interaction is actually a relativistic effect.

Hyperfine structure

Even closer inspection of the spectral lines with a very high resolution spectrometer reveals that the fine-structure lines are themselves split into more multiplets. The interactions that cause these splitting are called **hyperfine** interactions.

The hyperfine interactions are caused by the interactions between the electrons and the nucleus, as will be discussed in Section 7.7. The nuclear spin can interact with the magnetic field due to the orbital motion of the electron just as in spin-orbit coupling. This gives rise to shifts in the atomic energies that are about 2000 times smaller than the fine-structure shifts. The well-known 21 cm line of radio astronomy is caused by transitions between the hyperfine levels of atomic hydrogen. The photon energy in this case is 6×10^{-6} eV, or 0.05 cm^{-1}.

Chapter 2

Hydrogen

The quantum theory of hydrogen is the starting point for the whole subject of atomic physics. Bohr's derivation of the quantised energies was one of the triumphs of early quantum theory, and it is therefore helpful to begin the discussion of hydrogen by reviewing the Bohr model. We will then move on to the solution of the Schrödinger equation to find the allowed energies and angular momenta of the nucleus-electron system. In classical systems we are also able to calculate the precise trajectory of the orbit. This is not possible in quantum systems. The best we shall be able to do is to find the wave functions. These will then give us the probability amplitudes that allow us to calculate all the measurable properties of the system.

2.1 The Bohr model of hydrogen

The Bohr model of hydrogen is part of the "old" (i.e. pre-quantum mechanics) quantum theory of the atom. It includes the quantization of energy and angular momentum, but uses classical mechanics to describe the motion of the electron. With the advent of quantum mechanics, we realize that this is an inconsistent approach, and therefore should not be pushed too far. Nevertheless, the Bohr model does give the correct quantised energy levels of hydrogen, and also gives a useful parameter (the Bohr radius) for quantifying the size of atoms. Hence it remains a useful starting point for understanding the basic structure of atoms.

It is well known from classical physics that planetary orbits are characterized by their energy and angular momentum. We shall see that these are also key quantities in the quantum theory of the hydrogen atom. In 1911 Rutherford discovered the nucleus, which led to the idea of atoms consisting of electrons in classical orbits in which the central forces are provided by the Coulomb attraction to the positive nucleus, as shown in Fig. 2.1. The problem with this idea is that the electron in the orbit is constantly accelerating. Accelerating charges emit radiation called **bremsstrahlung**, and so the electrons should be radiating all the time. This would reduce the energy of the electron, and so it would gradually spiral into the nucleus, like an old satellite crashing to the earth. In 1913 Bohr resolved this issue by postulating that:

- The angular momentum L of the electron is quantized in units of $\hbar$ ($\hbar = h/2\pi$):

$$L = n\hbar\,, \tag{2.1}$$

  where n is an integer.

- The atomic orbits are stable, and light is only emitted or absorbed when the electron jumps from one orbit to another.

When Bohr made these hypotheses in 1913, they had no justification other than their success in predicting the energy spectrum of hydrogen. With hindsight, we realize that the first assumption is equivalent to stating that the circumference of the orbit must correspond to a fixed number of de Broglie wavelengths:

$$2\pi r = \text{integer} \times \lambda_{\rm deB} = n \times \frac{h}{p} = n \times \frac{h}{mv}\,, \tag{2.2}$$

which can be rearranged to give

$$L \equiv mvr = n \times \frac{h}{2\pi}\,. \tag{2.3}$$

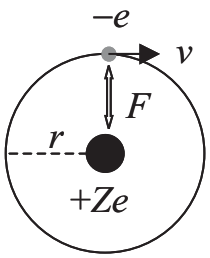

Figure 2.1: The Bohr model of the atom considers the electrons to be in orbit around the nucleus. The central force is provided by the Coulomb attraction. The angular momentum of the electron is quantized in integer units of $\hbar$.

The second assumption is a consequence of the fact that the Schrödinger equation leads to time-independent solutions (**eigenstates**).

The derivation of the quantized energy levels proceeds as follows. Consider an electron orbiting a nucleus of mass $m_{\rm N}$ and charge $+Ze$. The central force is provided by the Coulomb force:

$$F = \frac{mv^2}{r} = \frac{Ze^2}{4\pi\epsilon_0 r^2}\,. \tag{2.4}$$

As with all two-body orbit systems, the mass m that enters here is the **reduced mass** (see Appendix A):

$$\frac{1}{m} = \frac{1}{m_{\rm e}} + \frac{1}{m_{\rm N}}\,, \tag{2.5}$$

where $m_{\rm e}$ and $m_{\rm N}$ are the masses of the electron and the nucleus, respectively. The energy is given by:[1]

$$\begin{aligned} E_n &= \text{kinetic energy} + \text{potential energy} \\ &= \frac{1}{2}mv^2 - \frac{Ze^2}{4\pi\epsilon_0 r} \\ &= -\frac{mZ^2e^4}{8\epsilon_0^2 h^2 n^2}\,, \end{aligned} \tag{2.6}$$

where we made use of eqns 2.3 and 2.4 to solve for v and r. This can be written in the form:

$$E_n = -\frac{R'}{n^2} \tag{2.7}$$

where R' is given by:

$$R' = \left(\frac{m}{m_{\rm e}}Z^2\right) R_\infty hc\,, \tag{2.8}$$

and $R_\infty hc$ is the **Rydberg energy**:[2]

$$R_\infty hc = \frac{m_{\rm e}e^4}{8\epsilon_0^2 h^2}\,. \tag{2.9}$$

The Rydberg energy is a fundamental constant and has a value of 2.17987×10^{-18} J, which is equivalent to 13.606 eV. This tells us that the gross energy of the atomic states in hydrogen is of order 1 – 10 eV, or $10^4 - 10^5\,{\rm cm}^{-1}$ in wave number units.

R' is the effective Rydberg energy for the system in question. In the hydrogen atom we have an electron orbiting around a proton of mass $m_{\rm p}$. The reduced mass is therefore given by

$$m = m_{\rm e} \times \frac{m_{\rm p}}{m_{\rm e} + m_{\rm p}} = 0.9995\, m_{\rm e} \tag{2.10}$$

[1] In atoms the electron moves in free space, where the relative dielectric constant ϵ_r is equal to unity. However, in solid-state physics we frequently encounter hydrogenic systems inside crystals where ϵ_r is not equal to 1. In this case, we must replace ϵ_0 by $\epsilon_r\epsilon_0$ throughout.

[2] Note the difference between the Rydberg *energy* $R_\infty hc$ (13.606 eV) and the Rydberg *constant* R_∞ ($109{,}737\,{\rm cm}^{-1}$). The former has the dimensions of energy, while the latter has the dimensions of inverse length. They differ by a factor of hc. (See Table 2.1.) When high precision is not required, it is convenient just to use the symbol $R_{\rm H}$ for the Rydberg energy, although, strictly speaking, $R_{\rm H}$ differs from the true Rydberg energy by 0.05%. (See eqn 2.11.)

Quantity	Symbol	Formula	Numerical Value
Rydberg energy	$R_\infty hc$	$m_e e^4/8\epsilon_0^2 h^2$	2.17987×10^{-18} J 13.6057 eV
Rydberg constant	R_∞	$m_e e^4/8\epsilon_0^2 h^3 c$	109,737 cm^{-1}
Bohr radius	a_0	$\epsilon_0 h^2/\pi e^2 m_e$	5.29177×10^{-11} m
Fine structure constant	α	$e^2/2\epsilon_0 hc$	1/137.04

Table 2.1: Fundamental constants that arise from the Bohr model of the atom.

and the effective Rydberg energy for hydrogen is:

$$R_H = 0.9995\, R_\infty hc\,. \tag{2.11}$$

Atomic spectroscopy is very precise, and 0.05% factors such as this are easily measurable. Furthermore, in other systems such as positronium (an electron orbiting around a positron), the reduced mass effect is much larger, because $m = m_e/2$.

By following through the mathematics, we also find that the orbital radius and velocity are quantized. The relevant results are:

$$r_n = \frac{n^2}{Z}\,\frac{m_e}{m} a_0\,, \tag{2.12}$$

and

$$v_n = \alpha \frac{Z}{n}\, c\,. \tag{2.13}$$

The two fundamental constants that appear here are the **Bohr radius** a_0:

$$a_0 = \frac{h^2 \epsilon_0}{\pi m_e e^2}\,, \tag{2.14}$$

and the **fine structure constant** α:

$$\alpha = \frac{e^2}{2\epsilon_0 hc}\,. \tag{2.15}$$

The fundamental constants arising from the Bohr model are related to each other according to:

$$a_0 = \frac{\hbar}{m_e c}\frac{1}{\alpha}\,, \tag{2.16}$$

and

$$R_\infty hc = \frac{\hbar^2}{2m_e}\frac{1}{a_0^2}\,. \tag{2.17}$$

The definitions and values of these quantities are given in Table 2.1.

The energies of the photons emitted in transition between the quantized levels of hydrogen can be deduced from eqn 2.7:

$$h\nu = R_H \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)\,, \tag{2.18}$$

where n_1 and n_2 are the quantum numbers of the two states involved. Since $\nu = c/\lambda$, this can also be written in form:

$$\frac{1}{\lambda} = \frac{m}{m_e} R_\infty \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)\,. \tag{2.19}$$

In absorption we start from the ground state, so we put $n_1 = 1$. In emission, we can have any combination where $n_1 < n_2$. Some of the series of spectral lines have been given special names. The emission lines with $n_1 = 1$ are called the **Lyman series**, those with $n_1 = 2$ are called the **Balmer series**, etc. The Lyman and Balmer lines occur in the ultraviolet and visible spectral regions respectively.

A simple back-of-the-envelope calculation can easily show us that the Bohr model is not fully consistent with quantum mechanics. In the Bohr model, the linear momentum of the electron is given by:

$$p = mv = \left(\frac{\alpha Z}{n}\right) mc = \frac{n\hbar}{r_n} \,. \tag{2.20}$$

However, we know from the Heisenberg uncertainty principle that the precise value of the momentum must be uncertain. If we say that the uncertainty in the position of the electron is about equal to the radius of the orbit r_n, we find:

$$\Delta p \sim \frac{\hbar}{\Delta x} \approx \frac{\hbar}{r_n} \,. \tag{2.21}$$

On comparing Eqs. 2.20 and 2.21 we see that

$$|p| \approx n\Delta p \,. \tag{2.22}$$

This shows us that the magnitude of p is undefined except when n is large. This is hardly surprising, because the Bohr model is a mixture of classical and quantum models, and we can only expect the arguments to be fully self-consistent when we approach the classical limit at large n. For small values of n, the Bohr model fails when we take the full quantum nature of the electron into account.

2.2 The quantum mechanics of the hydrogen atom

The Bohr model presented in the previous section is only properly valid in the semiclassical limit. A fully consistent solution needs to use quantum mechanics throughout. Our task, therefore, is solve the Schrödinger equation for the hydrogen.

2.2.1 The Schrödinger Equation

The time-independent Schrödinger equation for hydrogen is given by:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{4\pi\epsilon_0 r}\right)\, \Psi(r,\theta,\phi) = E\, \Psi(r,\theta,\phi) \,, \tag{2.23}$$

where the spherical polar co-ordinates (r, θ, ϕ) refer to the position of the electron relative to the nucleus. Spherical polar co-ordinates are used here because the spherical symmetry of the atom facilitates the solution of the Schrödinger equation by the method of separation of variables. Since we are considering the motion of the electron relative to a stationary nucleus, the mass that appears in the Schrödinger equation is the **reduced mass** defined previously in eqn 2.5 and discussed in more detail in Appendix A. As we have already seen in eqn 2.10, the reduced mass of hydrogen has a value of $0.9995m_e$, which is very close to m_e.

Written out explicitly in spherical polar co-ordinates, the Schrödinger equation becomes:

$$-\frac{\hbar^2}{2m}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Psi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Psi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Psi}{\partial\phi^2}\right] - \frac{Ze^2}{4\pi\epsilon_0 r}\,\Psi = E\,\Psi \,. \tag{2.24}$$

Our task is to find the wave functions $\Psi(r, \theta, \phi)$ that satisfy this equation, and hence to find the allowed quantized energies E.

2.2.2 Separation of variables

The solution of the Schrödinger equation proceeds by the method of separation of variables. This works because the Coulomb potential is an example of a **central field** in which the force only lies along the radial direction. This allows us to separate the motion into the radial and angular parts:

$$\Psi(r,\theta,\phi) = R(r)\, F(\theta,\phi) \,. \tag{2.25}$$

We can re-write the Schrödinger equation in the following form:[3]

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Psi}{\partial r}\right) + \frac{\hat{\boldsymbol{L}}^2}{2mr^2}\Psi - \frac{Ze^2}{4\pi\epsilon_0 r}\Psi = E\,\Psi \,, \tag{2.26}$$

[3]Note that the "hat" symbol indicates that we are representing an operator and not just a number.

where the $\hat{\boldsymbol{L}}^2$ operator is:

$$\hat{\boldsymbol{L}}^2 = -\hbar^2 \left[\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right] . \tag{2.27}$$

This operator is derived from the **angular momentum operator** $\hat{\boldsymbol{L}}$. The properties of the angular momentum operator and the quantized angular momentum states of atoms will be considered in detail in Chapter 5. At this stage, we just consider a few basic points relating to the solution of the hydrogen atom.

On substituting eqn 2.25 into eqn 2.26, and noting that $\hat{\boldsymbol{L}}^2$ only acts on θ and ϕ, we find:

$$-\frac{\hbar^2}{2m} \frac{1}{r^2} \frac{\mathrm{d}}{\mathrm{d}r} \left(r^2 \frac{\mathrm{d}R}{\mathrm{d}r} \right) F + R \frac{\hat{\boldsymbol{L}}^2 F}{2mr^2} - \frac{Ze^2}{4\pi\epsilon_0 r} RF = E\, RF . \tag{2.28}$$

Multiply by r^2/RF and re-arrange to obtain:

$$-\frac{\hbar^2}{2m} \frac{1}{R} \frac{\mathrm{d}}{\mathrm{d}r} \left(r^2 \frac{\mathrm{d}R}{\mathrm{d}r} \right) - \frac{Ze^2 r}{4\pi\epsilon_0} - Er^2 = -\frac{1}{F} \frac{\hat{\boldsymbol{L}}^2 F}{2m} . \tag{2.29}$$

The left hand side is a function of r only, while the right hand side is only a function of the angular co-ordinates θ and ϕ. The only way this can be true is if both sides are equal to a constant. Let's call this constant $-\hbar^2\ell(\ell+1)/2m$, where ℓ is an arbitrary number that could be complex at this stage. This gives us, after a bit of re-arrangement:

$$-\frac{\hbar^2}{2m} \frac{1}{r^2} \frac{\mathrm{d}}{\mathrm{d}r} \left(r^2 \frac{\mathrm{d}R(r)}{\mathrm{d}r} \right) + \frac{\hbar^2\ell(\ell+1)}{2mr^2} R(r) - \frac{Ze^2}{4\pi\epsilon_0 r} R(r) = ER(r) , \tag{2.30}$$

and

$$\hat{\boldsymbol{L}}^2 F(\theta,\phi) = \hbar^2\ell(\ell+1) F(\theta,\phi) . \tag{2.31}$$

The task thus breaks down into one of solving two separate equations: one that describes the angular part of the wave function, and other dealing with the radial part.

2.2.3 The angular solution and the spherical harmonics

It is apparent from eqn 2.31 that the angular function $F(\theta,\phi)$ is an eigenfunction of the $\hat{\boldsymbol{L}}^2$ operator. These eigenfunctions are known as the **spherical harmonic** functions.

The spherical harmonics satisfy the equation:

$$\hat{\boldsymbol{L}}^2 Y(\theta,\phi) \equiv -\hbar^2 \left[\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial}{\partial\theta} \right) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right] Y(\theta,\phi) = L^2 Y(\theta,\phi) , \tag{2.32}$$

where L^2 is the eigenvalue of $\hat{\boldsymbol{L}}^2$. The solution of eqn 2.32 is considered in more detail in Appendix B. It turns out that solutions are only found in which L^2 takes the value $l(l+1)\hbar^2$, where l is 0 or a positive integer. l is called the **angular momentum quantum number**.

The spherical harmonics are also eigenfunctions of the operator that describes the z-component of the angular momentum, namely $\hat{L}_z$:

$$\hat{L}_z = -i\hbar \frac{\partial}{\partial\phi} . \tag{2.33}$$

The eigenvalue L_z is found by solving the equation:

$$\hat{L}_z Y(\theta,\phi) \equiv -i\hbar \frac{\partial Y}{\partial\phi} = L_z Y(\theta,\phi) . \tag{2.34}$$

Equation 2.34 implies that $Y(\theta,\phi) = f(\theta)\exp(-L_z\phi/i\hbar)$. The additional requirement that $Y(\theta,\phi)$ should be single-valued — i.e. $Y(\theta,\phi+2\pi) = Y(\theta,\phi)$ — implies that $L_z = m\hbar$, where m is an integer. m is called the **magnetic quantum number**, for reasons that will become apparent when we consider the effect of external magnetic fields in Chapter 8. Note that the same symbol m is used represent both the mass and the magnetic quantum number. Its meaning should be clear from the context, and, if necessary, we can add a subscript to the quantum number to distinguish it: m_l.

l	m	$Y_{lm}(\theta,\phi)$
0	0	$\sqrt{\frac{1}{4\pi}}$
1	0	$\sqrt{\frac{3}{4\pi}}\cos\theta$
1	±1	$\mp\sqrt{\frac{3}{8\pi}}\sin\theta e^{\pm i\phi}$
2	0	$\sqrt{\frac{5}{16\pi}}(3\cos^2\theta-1)$
2	±1	$\mp\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{\pm i\phi}$
2	±2	$\sqrt{\frac{15}{32\pi}}\sin^2\theta e^{\pm 2i\phi}$

Table 2.2: Spherical harmonic functions.

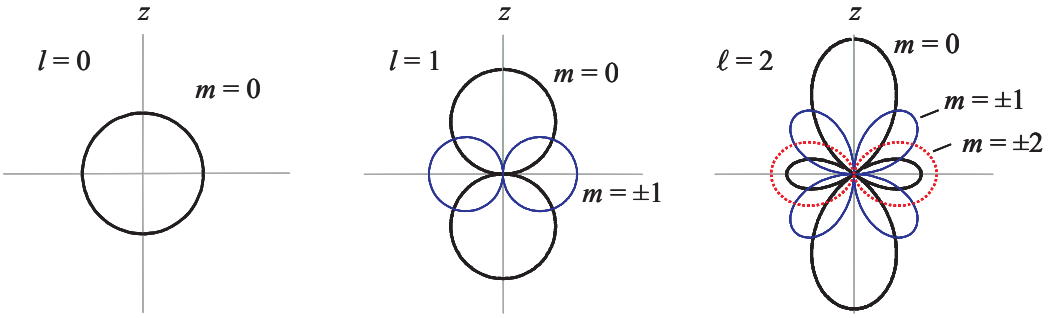

Figure 2.2: Polar plots of the spherical harmonics with $l \leq 2$. The plots are to be imagined with spherical symmetry about the z axis. In these polar plots, the value of the function for a given angle θ is plotted as the distance from the origin. Prettier pictures may be found, for example, at: `http://mathworld.wolfram.com/SphericalHarmonic.html`.

The final result is that the spherical harmonics are of the form:

$$Y_{lm}(\theta,\phi) = \text{normalization constant} \times P_l^m(\cos\theta)\,\mathrm{e}^{\mathrm{i}m\phi}\,, \tag{2.35}$$

where $P_l^m(\cos\theta)$ is a polynomial function in $\cos\theta$ called the associated Legendre polynominal, e.g. $P_0^0(\cos\theta) = \text{constant}$, $P_1^0(\cos\theta) = \cos\theta$, $P_1^{\pm1}(\cos\theta) = \sin\theta$, etc. The indices l and m must be integers, with $l \geq 0$ and $-l \leq m \leq +l$. In **spectroscopic notation**, states with $l = 0, 1, 2, 3, \ldots$ are called s, p, d, f, ... states, respectively.

The first few spherical harmonic functions are listed in Table 2.2. Representative polar plots of the wave functions are shown in figure 2.2. The spherical harmonics are **orthonormal** to each other, that is, they satisfy:

$$\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} Y_{lm}^*(\theta,\phi)Y_{l'm'}(\theta,\phi)\ \sin\theta\,\mathrm{d}\theta\mathrm{d}\phi = \delta_{l,l'}\delta_{m,m'}\,. \tag{2.36}$$

The symbol $\delta_{k,k'}$ is called the **Kronecker delta function**. It has the value of 1 if $k = k'$ and 0 if $k \neq k'$. The $\sin\theta$ factor in Eq. 2.36 comes from the volume increment in spherical polar co-ordinates: see Eq. 2.47 below.

On putting all this together, we see that the spherical harmonics (and hence the wave functions of the hydrogen atom) are eigenfunctions of both the $\hat{\boldsymbol{L}}^2$ and $\hat{L}_z$ operators:

$$\hat{\boldsymbol{L}}^2 Y_{lm}(\theta,\phi) = l(l+1)\hbar^2\, Y_{lm}(\theta,\phi)\,. \tag{2.37}$$

and

$$\hat{L}_z Y_{lm}(\theta,\phi) = m\hbar\, Y_{lm}(\theta,\phi)\,. \tag{2.38}$$

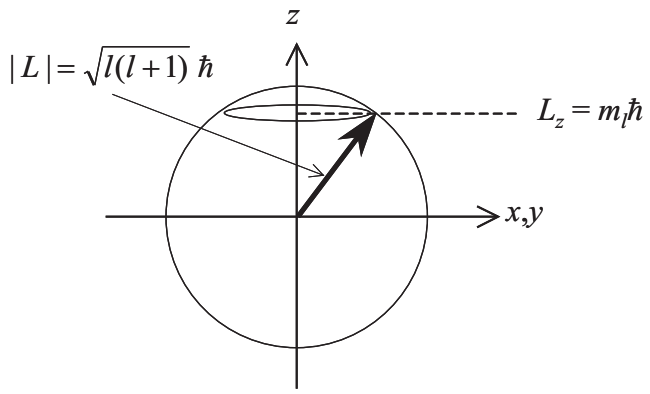

Figure 2.3: Vector model of the angular momentum in an atom. The angular momentum is represented by a vector of length $\sqrt{l(l+1)}\hbar$ precessing around the z-axis so that the z-component is equal to $m_l\hbar$.

On remembering that the allowed values of measurable quantities in quantum mechanics such as L^2 and L_z are found by solving eigenvalue equations, we can interpret eqns 2.37–2.38 as stating that the quantized states of the hydrogen atom have quantized angular momenta with magnitude equal to $\sqrt{l(l+1)}\hbar$ and a z-component component of $m\hbar$.[4] This is represented pictorially in the **vector model** of the atom shown in figure 2.3. In this model the angular momentum is represented as a vector of length $\sqrt{l(l+1)}\hbar$ angled in such a way that its component along the z axis is equal to $m\hbar$: we cannot specify the exact direction of $\boldsymbol{L}$, only $|\boldsymbol{L}|^2$ and L_z. As will be discussed in Section 5.2.1, the x and y components of the angular momentum are not known, because they do not commute with $\hat{L}_z$.

The quantisation of the magnitude of the angular momentum $|\boldsymbol{L}|^2$ with well-defined eigenvalues reflects the fact that the angular momentum of a classical particle interacting with a central field (i.e. one with a radial force parallel to $\boldsymbol{r}$) is a **constant of the motion**. This follows because the torque on the particle is zero, and so $\boldsymbol{L}$ must be a conserved quantity. (See discussion in Section 5.2.1.)

2.2.4 The radial wave functions and energies

We now return to the radial equation. On comparing eqns 2.37 and 2.31 we can now identify the arbitrary separation constant ℓ in the radial equation eqn 2.30 with the angular momentum quantum number l. On substituting $R(r) = P(r)/r$ into eqn 2.30 with $\ell = l$, we find:

$$\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{\hbar^2 l(l+1)}{2mr^2} - \frac{Ze^2}{4\pi\epsilon_0 r}\right] P(r) = EP(r)\,. \tag{2.39}$$

This now makes physical sense. It is a Schrödinger equation of the form:

$$\hat{H}P(r) = EP(r)\,, \tag{2.40}$$

where the energy operator $\hat{H}$ (i.e. the Hamiltonian) is given by:

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + V_{\text{effective}}(r)\,. \tag{2.41}$$

The first term in eqn 2.41 is the **radial kinetic energy** given by

$$\text{K.E.}_{\text{radial}} = \frac{p_r^2}{2m} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}r^2}\,.$$

The second term is the **effective potential energy**:

$$V_{\text{effective}}(r) = \frac{\hbar^2 l(l+1)}{2mr^2} - \frac{Ze^2}{4\pi\epsilon_0 r}\,, \tag{2.42}$$

[4] In Bohr's model, L was quantized in integer units of $\hbar$. (See eqn 2.3.) The full quantum treatment shows that this is only true in the classical limit where n is large and l approaches its maximum value, so that $L = \sqrt{l(l+1)}\hbar \sim \sqrt{(n-1)n}\hbar \sim n\hbar$.

Spectroscopic name	n	l	$R_{nl}(r)$
1s	1	0	$(Z/a_0)^{\frac{3}{2}} 2 \exp(-Zr/a_0)$
2s	2	0	$(Z/2a_0)^{\frac{3}{2}} 2 \left(1 - \frac{Zr}{2a_0}\right) \exp(-Zr/2a_0)$
2p	2	1	$(Z/2a_0)^{\frac{3}{2}} \frac{2}{\sqrt{3}} \left(\frac{Zr}{2a_0}\right) \exp(-Zr/2a_0)$
3s	3	0	$(Z/3a_0)^{\frac{3}{2}} 2 \left[1 - (2Zr/3a_0) + \frac{2}{3}\left(\frac{Zr}{3a_0}\right)^2\right] \exp(-Zr/3a_0)$
3p	3	1	$(Z/3a_0)^{\frac{3}{2}} (4\sqrt{2}/3) \left(\frac{Zr}{3a_0}\right) \left(1 - \frac{1}{2}\frac{Zr}{3a_0}\right) \exp(-Zr/3a_0)$
3d	3	2	$(Z/3a_0)^{\frac{3}{2}} (2\sqrt{2}/3\sqrt{5}) \left(\frac{Zr}{3a_0}\right)^2 \exp(-Zr/3a_0)$

Table 2.3: Radial wave functions of the hydrogen atom. a_0 is the Bohr radius (5.29×10^{-11} m). The wave functions are normalized so that $\int_{r=0}^{\infty} R_{nl}^* R_{nl} r^2 dr = 1$.

which has two components. The first of these is the orbital kinetic energy given by:

$$\text{K.E.}_{\text{orbital}} = \frac{L^2}{2I} = \frac{\hbar^2 l(l+1)}{2mr^2} \,,$$

where $I \equiv mr^2$ is the moment of inertia. The second is the usual potential energy due to the Coulomb energy.

This analysis shows that the quantized orbital motion adds quantized kinetic energy to the radial motion. For $l > 0$ the orbital kinetic energy will always be larger than the Coulomb energy at small r, and so the effective potential energy will be positive near $r = 0$. This has the effect of keeping the electron away from the nucleus, and explains why states with $l > 0$ have nodes at the origin (see below).

The wave function we require is given by Eq. 2.25. We have seen above that the $F(\theta, \phi)$ function that appears in Eq. 2.25 must be one of the spherical harmonics, some of which are listed in Table 2.2. The radial wave function $R(r)$ can be found by solving the radial differential equation given in Eq. 2.30 with $\ell = l$. The solution is given in section B.2 in Appendix B. The mathematics is somewhat complicated and here we just quote the main results.

Solutions are only found if we introduce an integer quantum number n. The energy depends only on n, but the functional form of $R(r)$ depends on both n and l, and so we must write the radial wave function as $R_{nl}(r)$. A list of some of the radial functions is given in Table 2.3, and representative wave functions are plotted in Fig. 2.4. The radial wave functions listed in Table 2.3 are of the form:

$$R_{nl}(r) = C_{nl} \cdot (\text{polynomial in } r) \cdot \mathrm{e}^{-r/a} \,, \tag{2.43}$$

where $a = na_{\mathrm{H}}/Z$, with a_{H} being the Bohr radius of Hydrogen given in eqn 2.14, namely 5.29×10^{-11} m. C_{nl} is a normalization constant. The polynomial functions that drop out of the equations are polynomials of order $n - 1$, and have $n - 1$ nodes. If $l = 0$, all the nodes occur at finite r, but if $l > 0$, one of the nodes is at $r = 0$.

The full wave function for hydrogen is therefore of the form:

$$\Psi_{nlm}(r, \theta, \phi) = R_{nl}(r) Y_{lm}(\theta, \phi) \,, \tag{2.44}$$

where $R_{nl}(r)$ is one of the radial functions given in eqn 2.43, and $Y_{lm}(\theta, \phi)$ is a spherical harmonic function as discussed in Section 2.2.3. The quantum numbers obey the following rules:

- n can have any integer value ≥ 1.
- l can have positive integer values from zero up to $(n - 1)$.
- m can have integer values from $-l$ to $+l$.

These rules drop out of the mathematical solutions. Functions that do not obey these rules will not satisfy the Schrödinger equation for the hydrogen atom.

The energy of the system is found to be:

$$E_n = -\frac{mZ^2e^4}{8\epsilon_0^2h^2}\frac{1}{n^2}\ , \tag{2.45}$$

which is the same as the Bohr formula given in Eq. 2.6. The energy only depends only on the principal quantum number n, which means that all the l states for a given value of n are **degenerate** (i.e. have the same energy), even though the radial wave functions depend on both n and l. This degeneracy with respect to l is called "accidental", and is a consequence of the fact that the electrostatic energy has a precise $1/r$ dependence in hydrogen. In more complex atoms, the electrostatic energy will depart from a pure $1/r$ dependence due to the shielding effect of inner electrons, and the gross energy will depend on l as well as n, even before we start thinking of higher-order fine-structure effects. We shall see an example of how this works in more detail when we consider the alkali atoms in section 4.5. Note also that the energy does not depend on the orbital quantum number m_l at all. Hence, the m_l states for each value of l are degenerate in the gross structure of all atoms in the absence of external fields.

The wave functions are nomalized so that

$$\int_{r=0}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\Psi^*_{n,l,m}\Psi_{n',l',m'}\ \mathrm{d}V = \delta_{n,n'}\delta_{l,l'}\delta_{m,m'} \tag{2.46}$$

where $\mathrm{d}V$ is the incremental volume element in spherical polar co-ordinates:

$$\mathrm{d}V = r^2\sin\theta\,\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi\,. \tag{2.47}$$

The radial probability function $P_{nl}(r)$ is the probability that the electron is found between r and $r+\mathrm{d}r$:

$$\begin{aligned} P_{nl}(r)\,\mathrm{d}r &= \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\Psi^*\Psi\ r^2\sin\theta\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi \\ &= |R_{nl}(r)|^2\,r^2\,\mathrm{d}r\ . \end{aligned} \tag{2.48}$$

The factor of r^2 that appears here is just related to the surface area of the radial shell of radius r (i.e. $4\pi r^2$.) Some representative radial probability functions are sketched in Fig. 2.5. 3-D plots of the shapes of the atomic orbitals are available at: `http://www.shef.ac.uk/chemistry/orbitron/`.

Expectation values of measurable quantities are calculated as follows:

$$\langle\hat{A}\rangle = \iiint\Psi^*\hat{A}\Psi\,\mathrm{d}V\ . \tag{2.49}$$

Thus, for example, the expectation value of the radius is given by:

$$\begin{aligned} \langle r\rangle &= \iiint\Psi^*r\Psi\mathrm{d}V \\ &= \int_{r=0}^{\infty}R^*_{nl}rR_{nl}r^2\mathrm{d}r\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}Y^*_{lm}(\theta,\phi)Y_{lm}(\theta,\phi)\ \sin\theta\mathrm{d}\theta\mathrm{d}\phi \\ &= \int_{r=0}^{\infty}R^*_{nl}rR_{nl}r^2\mathrm{d}r\,. \end{aligned} \tag{2.50}$$

This gives:

$$\langle r\rangle = \frac{n^2a_\mathrm{H}}{Z}\left(\frac{3}{2}-\frac{l(l+1)}{2n^2}\right)\ . \tag{2.51}$$

Note that this only approaches the Bohr value, namely n^2a_H/Z (see eqn 2.12), for the states with $l=n-1$ at large n.

2.3 Degeneracy and spin

We noted above that the m_l states of the hydrogen atom are all degenerate (i.e. have the same energy) in the absence of external fields. Furthermore, the **spin** of the electron does not appear in the basic Schrödinger equation for hydrogen given in eqn 2.24, which means that the energy does not depend on

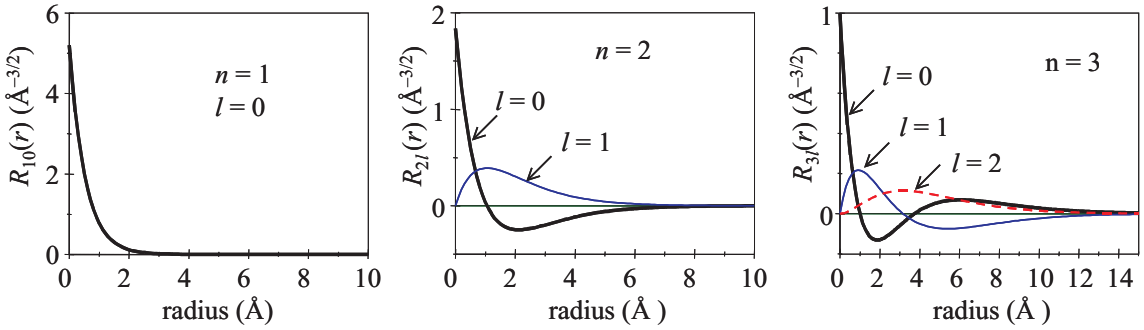

Figure 2.4: The radial wave functions $R_{nl}(r)$ for the hydrogen atom with $Z = 1$. Note that the axes for the three graphs are not the same.

l	Spectroscopic name	Degeneracy
0	s	2
1	p	6
2	d	10
3	f	14
⋮		⋮
l		$2(2l+1)$

Table 2.4: Degeneracy of the l states of the hydrogen atom.

the spin.[5] At this stage, we just note that electrons are spin 1/2 particles, with two states specified by the quantum number m_s, where $m_s = \pm 1/2$, for every quantized level. This means that each quantum state defined by the quantum numbers (n, l, m_l) has a degeneracy of two due to the two allowed spin states. Since each l state has $(2l+1)$ m_l levels, the full degeneracy of each l state is therefore $2\times(2l+1) = 2(2l+1)$, as listed in Table 2.4.

In hydrogen the l states are also degenerate. The degeneracy of the energy levels in hydrogen is therefore obtained by summing up the total number of all the (l, m_l, m_s) states that are possible for a given value of n:

$$\text{degeneracy} = 2 \times \sum_{l=0}^{n-1} (2l+1) = 2n^2 \,.$$

2.4 Hydrogen-like atoms

The theory of the hydrogen atom can be applied to any atom that consists of a single negative particle orbiting around a single positive one. There is a great variety of these **hydrogenic atoms**. They can treated by the same theory as developed here, but with the appropriate reduced mass included, and the appropriate value of Z. Here are some examples:

- Anti-hydrogen. This consists of a positron bound to an anti-proton. It should be exactly the same as hydrogen. Experiments are under way at CERN to make anti-hydrogen and measure the energy levels with very high precision. The discovery of a small difference in the spectra of hydrogen and anti-hydrogen might help to answer the question why there is no anti-matter in our known universe.

- Ionized atoms with $Z > 1$ in which all of the electrons have been stripped off apart from the last one, i.e. $A^{(Z-1)+}$. The simplest example is He^+, where $Z = 2$. We then have Li^{2+} ($Z = 3$), Be^{3+} ($Z = 4$), ..., etc. These would be written He II, Li III, Be IV, etc, in the spectroscopic notation explained in Section 1.2.

- Positronium. This consists of an electron bound to a positron. Since the positive particle has mass m_e, the reduced mass is $0.5m_e$. In solid-state physics, an exciton consists of an electron bound to a hole. The reduced mass is worked out from the effective masses of the electrons and holes, and the dielectric constant of the medium has to be included.

[5] The spin will eventually turn up in the Hamiltonian of hydrogen when we consider fine-structure effects.

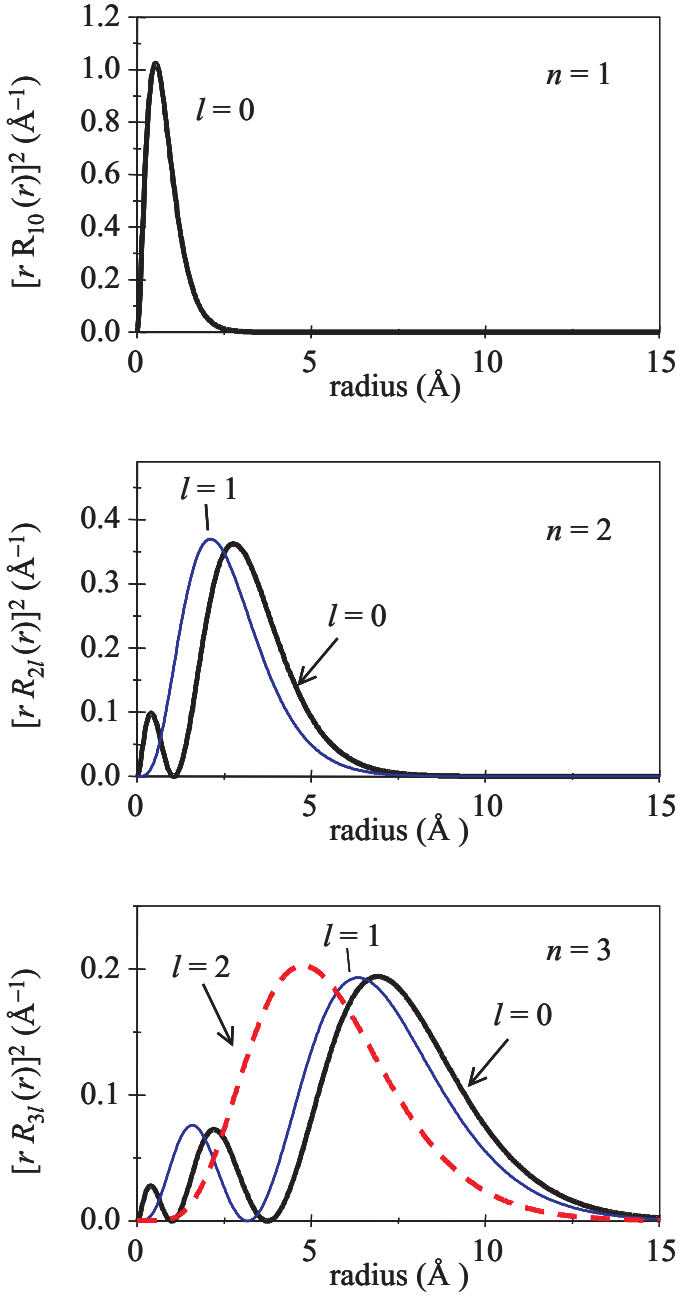

Figure 2.5: Radial probability functions for the first three n states of the hydrogen atom with $Z = 1$. Note that the radial probability is equal to $r^2|R_{nl}(r)|^2$, not just to $|R_{nl}(r)|^2$. Note also that the horizontal axes are the same for all three graphs, but not the vertical axes.

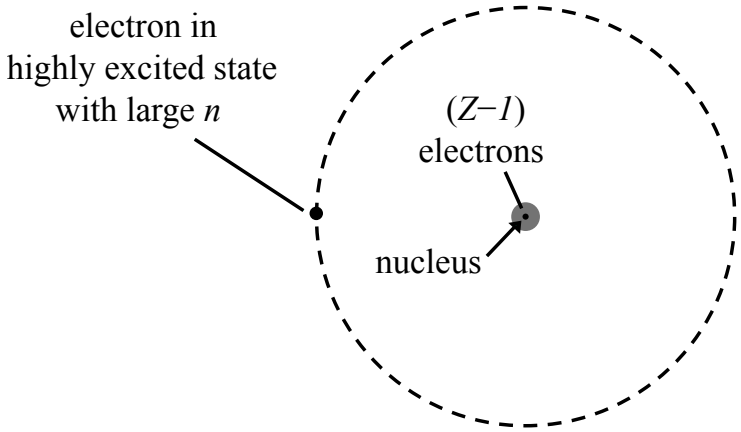

Figure 2.6: Rydberg atom. One of the electrons of a multi-electron atom is in a highly excited state far from the nucleus. The remaining $(Z-1)$ electrons are in tightly-bound states close to the nucleus.

- Impurity levels in semiconductors. These are modelled as electrons or holes bound to a positive or negative impurity atom. The impurity is bound to the crystal and therefore can be treated as having infinite mass. The effective mass must be used, and the dielectric constant of the medium.
- Muonium. This consists of an electron bound to a μ^+. The nucleus has mass $207m_e$, and hence $m = 0.995m_e$.
- Muonic hydrogen. This is a μ^- bound to a proton. The reduced mass is $186m_e$.

Another interesting application of hydrogen theory is in the study of **Rydberg atoms**. In this case, we study atoms in very highly excited states called Rydberg states, e.g. with $n \sim 100$. In the case of a neutral Rydberg atom with atomic number Z, there are $(Z-1)$ electrons in tightly-bound states close to the nucleus, and one electron in a very large radius state far from the nucleus, as shown in Fig. 2.6. The single outer electron has very low probability of overlapping the other electron wave functions, and so the central charge cloud close to the nucleus behaves as a net charge of $+e$, just as in hydrogen. The energies of the Rydberg states can thus be modelled as hydrogenic. With such large quantum numbers, the transition energies are in the microwave or radio-wave spectral regions. Since the radii are large, and the binding energies are small, the behaviour of Rydberg atoms is close to the semi-classical limit. Precision atomic spectroscopy can then test the convergence of classical and quantum theories in the limit of large n.

Reading

Demtröder, W., *Atoms, Molecules and Photons*, §3.4, 4.3 – §5.1.
Haken, H. and Wolf, H.C., *The Physics of Atoms and Quanta*, chapters 8 and 10.
Hertel, I.V. and Schulz, C.-P,, *Atoms, molecules and optical physics*, vol. 1 §1.8, §2.6
Phillips, A.C., *Introduction to Quantum Mechanics*, chapters 8 & 9.
Beisser, A., *Concepts of Modern Physics*, chapter 4 & 6.
Eisberg, R. and Resnick, R., *Quantum Physics*, chapter 7.

Chapter 3

Radiative transitions

In this chapter we shall look at the classical and quantum theories of radiative emission and absorption. This will enable us to derive certain selection rules which determine whether a particular transition is allowed or not. We shall also investigate the physical mechanisms that affect the shape of the spectral lines that are observed in atomic spectra.

3.1 Classical theories of radiating dipoles

The classical theories of radiation by atoms were developed at the end of the 19^{th} century before the discoveries of the electron and the nucleus. With the benefit of hindsight, we can understand more clearly how the classical theory works. We model the atom as a heavy nucleus with electrons attached to it by springs with different spring constants, as shown in Fig. 3.1(a). The spring represents the binding force between the nucleus and the electrons, and the values of the spring constants determine the resonant frequencies of each of the electrons in the atom. Every atom therefore has several different natural frequencies.

The nucleus is heavy, and so it does not move very easily at high frequencies. However, the electrons can readily vibrate about their mean position, as illustrated in Fig. 3.1(b). The vibrations of the electron create a fluctuating **electric dipole**. In general, electric dipoles consist of two opposite charges of $\pm q$ separated by a distance d. The dipole moment $\boldsymbol{p}$ is defined by:

$$\boldsymbol{p} = q\boldsymbol{d} \,, \tag{3.1}$$

where $\boldsymbol{d}$ is a vector of length d pointing from $-q$ to $+q$. In the case of atomic dipoles, the positive charge may be considered as being stationary, and so the time dependence of $\boldsymbol{p}$ is just determined by the movement of the electron:

$$p(t) = -ex(t) \,, \tag{3.2}$$

where $x(t)$ is the time dependence of the electron displacement.

It is well known that oscillating electric dipoles emit electromagnetic radiation at the oscillation frequency. This is how aerials work. Thus we expect an atom that has been excited into vibration to emit light waves at one of its natural resonant frequencies. This is the classical explanation of why atoms emit characteristic colours when excited electrically in a discharge tube. Furthermore, it is easy to see that an incoming light wave of frequency ω_0 can drive the natural vibrations of the atom through the oscillating force exerted on the electron by the electric field of the wave. This transfers energy from the light wave to the atom, which causes absorption at the resonant frequency. Hence the atom is also expected to absorb strongly at its natural frequency.

The classical theories actually have to assume that each electron has several natural frequencies of varying strengths in order to explain the observed spectra. If you do not do this, you end up predicting, for example, that hydrogen only has one emission frequency. There was no classical explanation of the origin of the atomic dipoles. It is therefore not surprising that we run into contradictions such as this when we try to patch up the model by applying our knowledge of electrons and nuclei gained by hindsight.

3.2 Quantum theory of radiative transitions

We have just seen that the classical model can explain why atoms emit and absorb light, but it does not offer any explanation for the frequency or the strength of the radiation. These can only be calculated

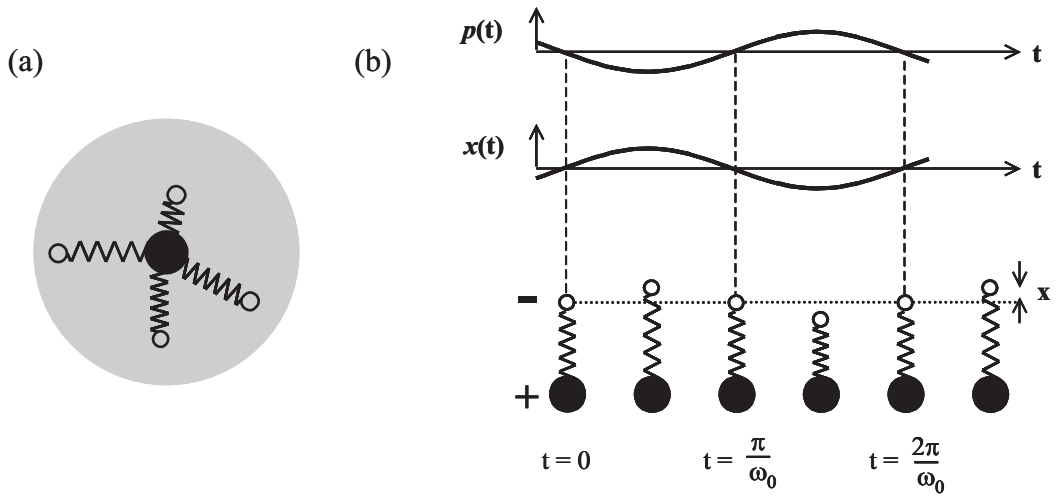

Figure 3.1: (a) Classical atoms consist of electrons bound to a heavy nucleus by springs with characteristic force constants. (b) The vibrations of an electron in an atom at its natural resonant frequency ω_0 creates an oscillating electric dipole. This acts like an aerial and emits electromagnetic waves at frequency ω_0. Alternatively, an incoming electromagnetic wave at frequency ω_0 can drive the oscillations at their resonant frequency. This transfers energy from the wave to the atom, which is equivalent to absorption.

by using quantum theory. Quantum theory tells us that atoms absorb or emit photons when they jump between quantized states, as shown in figure 3.2(a). The absorption or emission processes are called **radiative transitions**. The energy of the photon is equal to the difference in energy of the two levels:

$$h\nu = E_2 - E_1 \,. \tag{3.3}$$

Our task here is to calculate the *rate* at which these transitions occur.

The **transition rate** W_{12} can be calculated from the initial and final wave functions of the states involved by using **Fermi's golden rule**:

$$W_{12} = \frac{2\pi}{\hbar}|M_{12}|^2 g(h\nu)\,, \tag{3.4}$$

where M_{12} is the **matrix element** for the transition and $g(h\nu)$ is the **density of states**. The matrix element is equal to the overlap integral[1]:

$$M_{12} = \int \psi_2^*(\boldsymbol{r}) H'(\boldsymbol{r}) \psi_1(\boldsymbol{r})\, \mathrm{d}^3\boldsymbol{r}\,. \tag{3.5}$$

where H' is the **perturbation** that causes the transition. This represents the interaction between the atom and the light wave. There are a number of physical mechanisms that cause atoms to absorb or emit light. The strongest process is the electric dipole (E1) interaction. We therefore discuss E1 transitions first, leaving the discussion of higher order effects to Section 3.5.

The density of states factor is defined so that $g(h\nu)\mathrm{d}E$ is the number of *final* states per unit volume that fall within the energy range E to $E+\mathrm{d}E$, where $E = h\nu$. In the standard case of transitions between quantized levels in an atom, the initial and final electron states are *discrete*. In this case, the density of states factor that enters the golden rule is the density of *photon* states.[2] In free space, the photons can have any frequency and there is a continuum of states available, as illustrated in Fig. 3.2(b). The atom can therefore always emit a photon and it is the matrix element that determines the probability for this to occur. Hence we concentrate on the matrix element from now on.

[1] This is sometimes written in the shorthand **Dirac notation** as $M_{12} \equiv \langle 2|H'|1\rangle$.

[2] In solid-state physics, we consider transitions between electron *bands* rather than between discrete states. We then have to consider the density of electron states as well as the density of photon states when we calculate the transition rate. This point is covered in other courses, e.g. PHY475: Optical properties of solids.

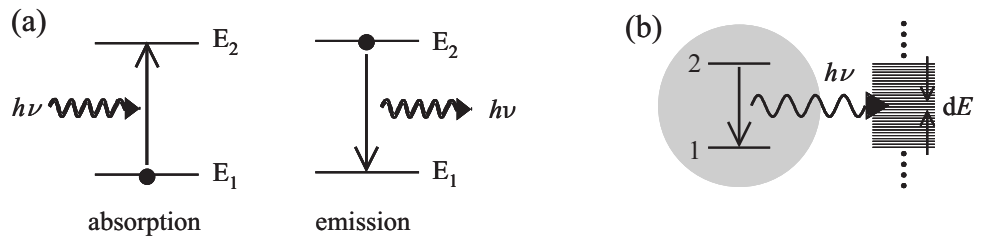

Figure 3.2: (a) Absorption and emission transitions in an atom. (b) Emission into a continuum of photon modes during a radiative transition between discrete atomic states.

3.3 Electric dipole (E1) transitions

Electric dipole transitions are the quantum mechanical equivalent of the classical dipole oscillator discussed in Section 3.1. We assume that the atom is irradiated with light, and makes a jump from level 1 to 2 by absorbing a photon. The interaction energy between an electric dipole $\boldsymbol{p}$ and an external electric field $\boldsymbol{\mathcal{E}}$ is given by

$$E = -\boldsymbol{p} \cdot \boldsymbol{\mathcal{E}} \,. \tag{3.6}$$

We presume that the nucleus is heavy, and so we only need to consider the effect on the electron. Hence the electric dipole perturbation is given by:

$$H' = +e\boldsymbol{r} \cdot \boldsymbol{\mathcal{E}} \,, \tag{3.7}$$

where $\boldsymbol{r}$ is the position vector of the electron and $\boldsymbol{\mathcal{E}}$ is the electric field of the light wave. This can be simplified to:

$$H' = e(x\mathcal{E}_x + y\mathcal{E}_y + z\mathcal{E}_z) \,, \tag{3.8}$$

where $\mathcal{E}_x$ is the component of the field amplitude along the x axis, etc. Now atoms are small compared to the wavelength of light, and so the amplitude of the electric field will not vary significantly over the dimensions of an atom. We can therefore take $\mathcal{E}_x$, $\mathcal{E}_y$, and $\mathcal{E}_z$ in Eq. 3.8 to be constants in the calculation, and just evaluate the following integrals:

$$\begin{aligned}
M_{12} &\propto \int \psi_1^* \, x \, \psi_2 \, \mathrm{d}^3\boldsymbol{r} \qquad x-\text{polarized light} \,, \\
M_{12} &\propto \int \psi_1^* \, y \, \psi_2 \, \mathrm{d}^3\boldsymbol{r} \qquad y-\text{polarized light} \,, \\
M_{12} &\propto \int \psi_1^* \, z \, \psi_2 \, \mathrm{d}^3\boldsymbol{r} \qquad z-\text{polarized light} \,.
\end{aligned} \tag{3.9}$$

Integrals of this type are called **dipole moments**. The dipole moment is thus the key parameter that determines the transition rate for the electric dipole process.

At this stage it is helpful to give a hand-waving explanation for why electric dipole transitions lead to the emission of light. To do this we need to to consider the time-dependence of the quantum mechanical wave functions. This naturally drops out of the *time-dependent* Schrödinger equation:

$$\hat{H}(\boldsymbol{r})\Psi(\boldsymbol{r},t) = \mathrm{i}\hbar\frac{\partial}{\partial t}\Psi(\boldsymbol{r},t) \,, \tag{3.10}$$

where $\hat{H}(\boldsymbol{r})$ is the Hamiltonian of the system. The solutions of Eq. 3.10 are of the form:

$$\Psi(\boldsymbol{r},t) = \psi(\boldsymbol{r})\mathrm{e}^{-\mathrm{i}Et/\hbar} \,, \tag{3.11}$$

where $\psi(\boldsymbol{r})$ satisfies the *time-independent* Schrödinger equation:

$$\hat{H}(\boldsymbol{r})\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}) \,. \tag{3.12}$$

During a transition between two quantum states of energies E_1 and E_2, the electron will be in a superposition state with a mixed wave function given by

$$\begin{aligned}
\Psi(\boldsymbol{r},t) &= c_1\Psi_1(\boldsymbol{r},t) + c_2\Psi_2(\boldsymbol{r},t) \\
&= c_1\psi_1(\boldsymbol{r})\mathrm{e}^{-\mathrm{i}E_1t/\hbar} + c_2\psi_2(\boldsymbol{r})\mathrm{e}^{-\mathrm{i}E_2t/\hbar} \,,
\end{aligned} \tag{3.13}$$

Quantum number	Selection rule	
parity	changes	
l	$\Delta l = \pm 1$	
m	$\Delta m = 0, \pm 1$	unpolarized light
	$\Delta m = 0$	linear polarization $\parallel z$
	$\Delta m = \pm 1$	linear polarization in (x, y) plane
	$\Delta m = +1$	σ^+ circular polarization
	$\Delta m = -1$	σ^- circular polarization
s	$\Delta s = 0$	
m_s	$\Delta m_s = 0$	

Table 3.1: Electric dipole selection rules for the quantum numbers of the states involved in the transition.

where c_1 and c_2 are the mixing coefficients. The expectation value $\langle x \rangle$ of the position of the electron is given by:

$$\langle x \rangle = \int \Psi^* \, x \, \Psi \, \mathrm{d}^3\boldsymbol{r} \, . \tag{3.14}$$

With Ψ given by Eq. 3.13 we obtain:

$$\begin{aligned} \langle x \rangle &= c_1^* c_1 \int \psi_1^* \, x \, \psi_1 \, \mathrm{d}^3\boldsymbol{r} + c_2^* c_2 \int \psi_2^* \, x \, \psi_2 \, \mathrm{d}^3\boldsymbol{r} \\ &+ c_1^* c_2 \mathrm{e}^{-\mathrm{i}(E_2 - E_1)t/\hbar} \int \psi_1^* \, x \, \psi_2 \, \mathrm{d}^3\boldsymbol{r} \; + \; c_2^* c_1 \mathrm{e}^{-\mathrm{i}(E_1 - E_2)t/\hbar} \int \psi_2^* \, x \, \psi_1 \, \mathrm{d}^3\boldsymbol{r} \, . \end{aligned} \tag{3.15}$$

This shows that if the dipole moment defined in Eq. 3.9 is non-zero, then the electron wave-packet oscillates in space at angular frequency $(E_2 - E_1)/\hbar$. The oscillation of the electron wave packet creates an oscillating electric dipole, which then radiates light at angular frequency $(E_2 - E_1)/\hbar$. Hey presto!

3.4 Selection rules for E1 transitions

Electric dipole transitions can only occur if the **selection rules** summarized in Table 3.1 are satisfied. Transitions that obey these E1 selection rules are called **allowed** transitions. If the selection rules are not satisfied, the matrix element (i.e. the dipole moment) is zero, and we then see from Eq. 3.4 that the transition rate is zero. The origins of these rules are discussed below.

Parity

The parity of a function refers to the sign change under inversion about the origin. Thus if $f(-x) = f(x)$ we have even parity, whereas if $f(-x) = -f(x)$ we have odd parity. Now atoms are spherically symmetric, which implies that

$$|\psi(-\boldsymbol{r})|^2 = |\psi(+\boldsymbol{r})|^2 \, . \tag{3.16}$$

Hence we must have that

$$\psi(-\boldsymbol{r}) = \pm\psi(+\boldsymbol{r}) \, . \tag{3.17}$$

In other words, the wave functions have either even or odd parity. The dipole moment of the transition is given by Eq. 3.9. x, y and z are odd functions, and so the product $\psi_1\psi_2$ must be an odd function if M_{12} is to be non-zero. Hence ψ_1 and ψ_2 must have different parities.

The orbital quantum number l

The parity of the spherical harmonic functions is equal to $(-1)^l$. Hence the parity selection rule implies that Δl must be an odd number. Detailed evaluation of the overlap integrals tightens this rule to $\Delta l = \pm 1$. This can be seen as a consequence of the fact that the angular momentum of a photon is $\pm\hbar$, with the sign depending on whether we have a left or right circularly polarized photon. Conservation of angular momentum therefore requires that the angular momentum of the atom must change by one unit.

The magnetic quantum number m

The dipole moment for the transition can be written out explicitly:

$$M_{12} \propto \int_{r=0}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \Psi^*_{n',l',m'}\, \boldsymbol{r}\, \Psi_{n,l,m}\, r^2 \sin\theta \,\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi\,. \tag{3.18}$$

We consider here just the ϕ part of this integral:

$$M_{12} \propto \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}m'\phi}\, \boldsymbol{r}\, \mathrm{e}^{\mathrm{i}m\phi}\,\mathrm{d}\phi\,, \tag{3.19}$$

where we have made use of the fact that (see eqns 2.44 and 2.35):

$$\Psi_{n,l,m}(r,\theta,\phi) \propto \mathrm{e}^{\mathrm{i}m\phi}\,. \tag{3.20}$$

Now for z-polarized light we have from Eq. 3.9:

$$M_{12} \propto \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}m'\phi}\, z\, \mathrm{e}^{\mathrm{i}m\phi}\,\mathrm{d}\phi \propto \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}m'\phi}\cdot 1 \cdot \mathrm{e}^{\mathrm{i}m\phi}\,\mathrm{d}\phi\,, \tag{3.21}$$

because $z = r\cos\theta$. Hence we must have that $m' = m$ if M_{12} is to be non-zero. If the light is polarized in the (x, y) plane, we have integrals like

$$M_{12} \propto \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}m'\phi}\, x\, \mathrm{e}^{\mathrm{i}m\phi}\,\mathrm{d}\phi \propto \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}m'\phi}\cdot \mathrm{e}^{\pm\mathrm{i}\phi} \cdot \mathrm{e}^{\mathrm{i}m\phi}\,\mathrm{d}\phi\,. \tag{3.22}$$

This is because $x = r\sin\theta\cos\phi = r\sin\theta\frac{1}{2}(\mathrm{e}^{+\mathrm{i}\phi} + \mathrm{e}^{-\mathrm{i}\phi})$, and similarly for y. This give $m' - m = \pm 1$. This rule can be tightened up a bit by saying that $\Delta m = +1$ for σ^+ circularly polarized light and $\Delta m = -1$ for σ^- circularly polarized light. If the light is unpolarized, then all three linear polarizations are possible, and we can have $\Delta m = 0, \pm 1$.

Spin

The photon does not interact with the electron spin. Therefore, the spin state of the atom does not change during the transition. This implies that the spin quantum numbers s and m_s are unchanged.

3.5 Higher order transitions

How does an atom de-excite if E1 transitions are forbidden by the selection rules? In some cases it may be possible for the atom to de-excite by alternative methods. For example, the 3s $\rightarrow$ 1s transition is forbidden, but the atom can easily de-excite by two allowed E1 transitions, namely 3s $\rightarrow$ 2p, then 2p $\rightarrow$ 1s. However, this may not always be possible, and in these cases the atom must de-excite by making a **forbidden** transition. The use of the word "forbidden" is somewhat misleading here. It really means "electric-dipole forbidden". The transitions are perfectly possible, but they just occur at a slower rate.

After the electric-dipole interaction, the next two strongest interactions between the photon and the atom give rise to **magnetic dipole** (M1) and **electric quadrupole** (E2) transitions. There have different selection rules to E1 transitions (e.g. parity is conserved), and may therefore be allowed when when E1 transitions are forbidden. M1 and E2 transitions are second-order processes and have much smaller probabilities than E1 transitions.

In extreme cases it may happen that all types of radiative transitions are forbidden. In this case, the excited state is said to be **metastable**, and must de-excite by transferring its energy to other atoms in collisional processes or by multi-photon emission.

3.6 Radiative lifetimes

An atom in an excited state has a spontaneous tendency to de-excite by a radiative transition involving the emission of a photon. This follows from statistical physics: atoms with excess energy tend to want to get rid of it. This process is called **spontaneous emission**. Let us suppose that there are N_2 atoms

Transition	Einstein A coefficient	Radiative lifetime
E1 allowed	$10^8 - 10^9$ s^{-1}	1 – 10 ns
E1 forbidden (M1 or E2)	$10^3 - 10^6$ s^{-1}	1 μs – 1 ms

Table 3.2: Typical transition rates and radiative lifetimes for allowed and forbidden transitions at optical frequencies.

in level 2 at time t. We use quantum mechanics to calculate the transition rate from level 2 to level 1, and then write down a rate equation for N_2 as follows:

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -AN_2\,. \tag{3.23}$$

This merely says that the total number of atoms making transitions is proportional to the number of atoms in the excited state and to the quantum mechanical probability. The parameter A that appears in eqn 3.23 is called the **Einstein A coefficient** of the transition. The Einstein B coefficients that describe the processes of stimulated emission and absorption are considered in Section 9.3 in the context of laser physics.

Equation 3.23 has the following solution:

$$\begin{aligned} N_2(t) &= N_2(0)\exp(-At) \\ &= N_2(0)\exp(-t/\tau)\,, \end{aligned} \tag{3.24}$$

where

$$\tau = \frac{1}{A}\,. \tag{3.25}$$

Equation 3.24 shows that if the atoms are excited into the upper level, the population will decay due to spontaneous emission with a time constant τ. τ is thus called the **natural radiative lifetime** of the excited state.

The values of the Einstein A coefficient and hence the radiative lifetime τ vary considerably from transition to transition. Allowed E1 transitions have A coefficients in the range $10^8 - 10^9\,\mathrm{s}^{-1}$ at optical frequencies, giving radiative lifetimes of $\sim 1 - 10$ ns. Forbidden transitions, on the other hand, are much slower because they are higher order processes. The radiative lifetimes for M1 and E2 transitions are typically in the millisecond or microsecond range. This point is summarized in Table 3.2.

3.7 The width and shape of spectral lines

The radiation emitted in atomic transitions is not perfectly monochromatic. The shape of the emission line is described by the **spectral line shape function** $g(\nu)$. This is a function that peaks at the line centre defined by

$$h\nu_0 = (E_2 - E_1)\,, \tag{3.26}$$

and is normalized so that:

$$\int_0^\infty g(\nu)\,\mathrm{d}\nu = 1\,. \tag{3.27}$$

The most important parameter of the line shape function is the **full width at half maximum** (FWHM) $\Delta\nu$, which quantifies the width of the spectral line. We shall see below how the different types of line broadening mechanisms give rise to two common line shape functions, namely the **Lorentzian** and **Gaussian** functions.

In a gas of atoms, spectral lines are broadened by three main processes:

- natural broadening,
- collision broadening,
- Doppler broadening.

We shall look at each of these processes separately below. A useful general division can be made at this stage by classifying the broadening as either **homogeneous** or **inhomogeneous**.

- **Homogeneous** broadening affects all the individual atoms in the same way. Natural lifetime and collision broadening are examples of homogeneous processes. All the atoms are behaving in the same way, and each atom produces the same emission spectrum.
- **Inhomogeneous** broadening affects different individual atoms in different ways. Doppler broadening is the standard example of an inhomogeneous process. The individual atoms are presumed to behave identically, but they are moving at different velocities, and one can associate different parts of the spectrum with the subset of atoms with the appropriate velocity. Inhomogeneous broadening is also found in solids, where different atoms may experience different local environments due to the inhomogeneity of the medium.

3.8 Natural broadening

We have seen in Section 3.6 that the process of spontaneous emission causes the excited states of an atom to have a finite lifetime. Let us suppose that we somehow excite a number of atoms into level 2 at time $t = 0$. Equation 3.23 shows us that the rate of transitions is proportional to the instantaneous population of the upper level, and eqn 3.24 shows that this population decays exponentially. Thus the rate of atomic transitions decays exponentially with time constant τ. For every transition from level 2 to level 1, a photon of angular frequency $\omega_0 = (E_2 - E_1)/\hbar$ is emitted. Therefore a burst of light with an exponentially-decaying intensity will be emitted for $t > 0$:

$$I(t) = I(0) \exp(-t/\tau) \,. \tag{3.28}$$

This corresponds to a time dependent electric field of the form:

$$\begin{aligned} t < 0 : \qquad \mathcal{E}(t) &= 0 \,, \\ t \geq 0 : \qquad \mathcal{E}(t) &= \mathcal{E}_0 \, \mathrm{e}^{\mathrm{i}\omega_0 t} \, \mathrm{e}^{-t/2\tau} \,. \end{aligned} \tag{3.29}$$

The extra factor of 2 in the exponential in eqn 3.29 compared to eqn 3.28 arises because $I(t) \propto \mathcal{E}(t)^2$. We now take the Fourier transform of the electric field to derive the frequency spectrum of the burst:

$$\mathcal{E}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathcal{E}(t) \, \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t \,. \tag{3.30}$$

The emission spectrum is then given by:

$$I(\omega) \propto \left| \mathcal{E}(\omega)^2 \right| \propto \frac{1}{(\omega - \omega_0)^2 + (1/2\tau)^2} \,. \tag{3.31}$$

Remembering that $\omega = 2\pi\nu$, we find the final result for the spectral line shape function:

$$g(\nu) = \frac{\Delta\nu}{2\pi} \, \frac{1}{(\nu - \nu_0)^2 + (\Delta\nu/2)^2} \,, \tag{3.32}$$

where the full width at half maximum is given by

$$\Delta\nu = \frac{1}{2\pi\tau} \,. \tag{3.33}$$

The spectrum described by eqn 3.32 is called a **Lorentzian** line shape. This function is plotted in Fig. 3.3. Note that we can re-write eqn 3.33 in the following form:

$$\Delta\nu \cdot \tau = \frac{1}{2\pi} \,. \tag{3.34}$$

By multiplying both sides by h, we can recast this as:

$$\Delta E \cdot \tau = h/2\pi \,. \tag{3.35}$$

If we realize that τ represents the average time the atom stays in the excited state (i.e the uncertainty in the time), we can interpret this as the **energy–time uncertainty principle**.

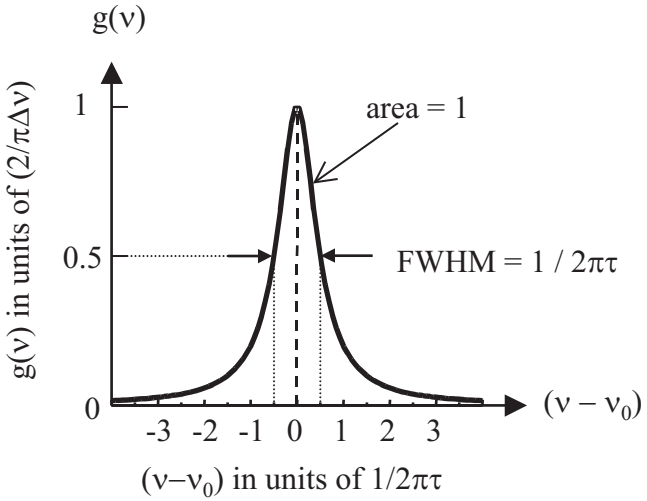

Figure 3.3: The Lorentzian line shape. The functional form is given in eqn 3.32. The function peaks at the line centre ν_0 and has an FWHM of $1/2\pi\tau$. The function is normalized so that the total area is unity.

3.9 Collision (Pressure) broadening

The atoms in a gas jostle around randomly and frequently collide into each other and the walls of the containing vessel. This interrupts the process of light emission and effectively shortens the lifetime of the excited state. This gives additional line broadening through the uncertainty principle, as determined by eqn 3.33 with τ replaced by τ_c, where τ_c is the mean time between collisions.

It can be shown from the kinetic theory of gases that the time between collisions in an ideal gas is given by:

$$\tau_c \sim \frac{1}{\sigma_s P}\left(\frac{\pi m k_B T}{8}\right)^{1/2}, \tag{3.36}$$

where σ_s is the collision cross-section, and P is the pressure. The collision cross-section is an effective area which determines whether two atoms will collide or not. It will be approximately equal to the size of the atom. For example, for sodium atoms we have:

$$\sigma_s \sim \pi r_{atom}^2 \sim \pi \times (0.2\,\mathrm{nm})^2 = 1.2 \times 10^{-19}\,\mathrm{m}^2\,.$$

Thus at S.T.P we find $\tau_c \sim 6 \times 10^{-10}$ s, which gives a line width of ~ 1 GHz. Note that τ_c is much shorter than typical radiative lifetimes. For example, the strong yellow D-lines in sodium have a radiative lifetime of 16 ns, which is nearly two orders of magnitude larger.

In conventional atomic discharge tubes, we reduce the effects of pressure broadening by working at low pressures. We see from eqn 3.36 that this increases τ_c, and hence reduces the linewidth. This is why we tend to use "low pressure" discharge lamps for spectroscopy.

3.10 Doppler broadening

The spectrum emitted by a typical gas of atoms in a low pressure discharge lamp is usually found to be much broader than the radiative lifetime would suggest, even when everything is done to avoid collisions. For example, the radiative lifetime for the 632.8 nm line in neon is 2.7×10^{-7} s. Equation 3.33 tells us that we should have a spectral width of 0.54 MHz. In fact, the line is about three orders of magnitude broader, and moreover, does not have the Lorentzian lineshape given by eqn 3.32.

The reason for this discrepancy is the thermal motion of the atoms. The atoms in a gas move about randomly with a root-mean-square thermal velocity given by:

$$\tfrac{1}{2} m v_x^2 = \tfrac{1}{2} k_B T\,, \tag{3.37}$$

where k_B is Boltzmann's constant. At room temperature the thermal velocities are quite large. For example, for sodium with a mass number of 23 we find $v_x \sim 330\,\mathrm{ms}^{-1}$ at 300 K. This random thermal

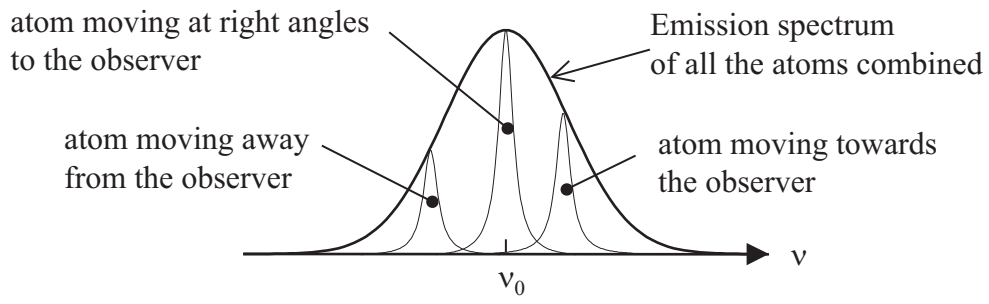

Figure 3.4: The Doppler broadening mechanism. The thermal motion of the atoms causes their frequency to be shifted by the Doppler effect.

motion of the atoms gives rise to Doppler shifts in the observed frequencies, which then cause line broadening, as illustrated in Fig. 3.4. This is Doppler line broadening mechanism.

Let us suppose that the atom is emitting light from a transition with centre frequency ν_0. An atom moving with velocity v_x will have its observed frequency shifted by the Doppler effect according to:

$$\nu = \nu_0 \left(1 \pm \frac{v_x}{c}\right) , \tag{3.38}$$

where the + and − sign apply to motion towards or away from the observer respectively. The probability that an atom has velocity v_x is governed by the Boltzmann formula:

$$p(E) \propto \mathrm{e}^{-E/k_\mathrm{B}T} . \tag{3.39}$$

On setting E equal to the kinetic energy, we find that the number of atoms with velocity v_x is given by the Maxwell–Boltzmann distribution:

$$N(v_x) \propto \exp\left(-\frac{\frac{1}{2}mv_x^2}{k_\mathrm{B}T}\right) . \tag{3.40}$$

We can combine eqns 3.38 and 3.40, to find the number of atoms emitting at frequency ν:

$$N(\nu) \propto \exp\left(-\frac{mc^2(\nu-\nu_0)^2}{2k_\mathrm{B}T\nu_0^2}\right) . \tag{3.41}$$

The frequency dependence of the light emitted is therefore given by:

$$I(\nu) \propto \exp\left(-\frac{mc^2(\nu-\nu_0)^2}{2k_\mathrm{B}T\nu_0^2}\right) . \tag{3.42}$$

This gives rise to a **Gaussian** line shape with $g(\nu)$ given by:

$$g(\nu) \propto \exp\left(-\frac{mc^2(\nu-\nu_0)^2}{2k_\mathrm{B}T\nu_0^2}\right) , \tag{3.43}$$

with a full width at half maximum equal to:

$$\Delta\nu_\mathrm{D} = 2\nu_0\left(\frac{(2\ln 2)k_\mathrm{B}T}{mc^2}\right)^{1/2} = \frac{2}{\lambda}\left(\frac{(2\ln 2)k_\mathrm{B}T}{m}\right)^{1/2} . \tag{3.44}$$

The Doppler linewidth in a gas at S.T.P is usually several orders of magnitude larger than the natural linewidth. For example, the Doppler line width of the 632.8 nm line of neon at 300 K works out to be 1.3 GHz, i.e. three orders of magnitude larger than the broadening due to spontaneous emission. The dominant broadening mechanism in the emission spectrum of gases at room temperature is usually Doppler broadening, and the line shape is closer to Gaussian than Lorentzian. [3]

[3]Since $\Delta\nu_\mathrm{D}$ is proportional to $\sqrt{T}$, we can reduce its value by cooling the gas. Cooling also reduces the collision broadening because $P \propto T$, and therefore $\tau_\mathrm{c} \propto T^{-1/2}$. (See eqn 3.36.) Laser cooling techniques can produce temperatures in the micro-Kelvin range, where we finally observe the natural line shape of the emission line.

3.11 Converting between line widths in frequency and wavelength units

Spectral lines can be plotted against frequency, photon energy, wave number or wavelength. Converting between line widths for the first three of these presents no difficulty, since it just involves a linear scaling. (See Section 1.5.) However, converting to wavelengths is more complicated, because of the inverse relationship between wavelength and frequency.

Let us suppose that we have an atomic transition of centre frequency ν_0 and FWHM $\Delta\nu$, where $\Delta\nu \ll \nu_0$. We convert to wavelengths through $\nu = c/\lambda$. This implies that:

$$\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = -\frac{c}{\lambda^2}\,, \tag{3.45}$$

and hence that the FWHM in wavelength units is given by:

$$\Delta\lambda = \left|-\frac{\lambda_0^2}{c}\Delta\nu\right| = \frac{\lambda_0^2}{c}\Delta\nu\,, \tag{3.46}$$

where $\lambda_0 = c/\nu_0$. A simple way of remembering this follows directly from eqn 3.46, namely:

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta\nu}{\nu}\,, \tag{3.47}$$

where we have dropped the subscripts on the centre frequency and wavelength.

Equations 3.46 and 3.47 work in the limit where $\Delta\nu \ll \nu_0$, or equivalently, $\Delta\lambda \ll \lambda_0$. In some cases (e.g. in molecular physics or solid-state physics) we might be considering a broad emission band rather than a narrow spectral line. In this situation, we have to go back to first principles to convert between frequency and wavelength units. Suppose that the emission band runs from frequency ν_1 to ν_2. The spectral width in wavelength units is then worked out from:

$$\Delta\lambda = |\lambda_2 - \lambda_1| = \left|\frac{c}{\nu_2} - \frac{c}{\nu_1}\right|\,. \tag{3.48}$$

Here, as in eqn 3.46, the modulus is needed because an increase in frequency causes a decrease in wavelength, and vice versa. Note that eqn 3.48 always works, and can be applied to the case of narrow spectral lines by putting $\nu_1 = \nu_0 - \Delta\nu/2$ and $\nu_2 = \nu_0 + \Delta\nu/2$, or, more easily, $\nu_1 = \nu_0$ and $\nu_2 = \nu_0 + \Delta\nu$.

3.12 Atoms in solids

In laser physics we shall frequently be interested in the emission spectra of atoms in crystals. The spectra will be subject to lifetime broadening as in gases, since this is a fundamental property of radiative emission. However, the atoms are locked in a lattice, and so collisional broadening is not relevant. Doppler broadening does not occur either, for the same reason. On the other hand, the emission lines can be broadened by other mechanisms.

In some cases it may be possible for the atoms to de-excite from the upper level to the lower level by making a **non-radiative transition**. One way this could happen is to drop to the lower level by emitting phonons (ie heat) instead of photons. To allow for this possibility, we must re-write eqn 3.23 in the following form:

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -AN_2 - \frac{N_2}{\tau_{\mathrm{NR}}} = -\left(A + \frac{1}{\tau_{\mathrm{NR}}}\right)N_2 = -\frac{N_2}{\tau}\,, \tag{3.49}$$

where τ_{NR} is the non-radiative transition time. This shows that non-radiative transitions shorten the lifetime of the excited state according to:

$$\frac{1}{\tau} = A + \frac{1}{\tau_{\mathrm{NR}}}\,. \tag{3.50}$$

We thus expect additional lifetime broadening according to eqn 3.33. The phonon emission times in solids are often very fast, and can cause substantial broadening of the emission lines. This is the solid-state equivalent of collisional broadening.

Another factor that may cause line broadening is the inhomogeneity of the host medium, for example when the atoms are doped into a glass. If the environment in which the atoms find themselves is not entirely uniform, the emission spectrum will be affected through the interaction between the atom and the local environment. This is an example of an inhomogeneous broadening mechanism.

Reading

Bransden and Joachain, *Physics of Atoms and Molecules*, chapter 4
Demtröder, W., *Atoms, Molecules and Photons*, §7.1 – §7.4.
Haken, H. and Wolf, H.C., *The Physics of Atoms and Quanta*, chapter 16.
Hertel and Schulz, *Atoms, Molecules and Optical Physics, 1*, §4.2 – 4.4, 5.1
Hooker, S. and Webb, C., *Laser Physics*, chapter 3.
Smith, F.G. and King, T.A., *Optics and Photonics*, sections 13.1–4, 20.1–2
Beisser, A., *Concepts of Modern Physics*, sections 6.8–9
Eisberg, R. and Resnick, R., *Quantum Physics*, section 8.7.

Chapter 4

The shell model and alkali spectra

Everything we have been doing so far in this course applies to hydrogenic atoms. We have taken this approach because the hydrogen atom only contains two particles: the nucleus and the electron. This is a **two-body system** and can be solved exactly by separating the motion into the centre of mass and relative co-ordinates. This has allowed us to find the wave functions and understand the meaning of the quantum numbers n, l, m_l and m_s.

We are well aware, however, that hydrogen is only the first of about 100 elements. These are not two body problems: we have one nucleus and many electrons, which is a **many body problem**, with no exact solution. This chapter begins our consideration of the approximation techniques that are used to understand the behaviour of many-electron atoms.

4.1 The central field approximation

The Hamiltonian for an N-electron atom with nuclear charge $+Ze$ can be written in the form:

$$\hat{H} = \sum_{i=1}^{N} \left(-\frac{\hbar^2}{2m} \boldsymbol{\nabla}_i^2 - \frac{Ze^2}{4\pi\epsilon_0 r_i} \right) + \sum_{i>j}^{N} \frac{e^2}{4\pi\epsilon_0 r_{ij}} \,, \tag{4.1}$$

where $N = Z$ for a neutral atom. The subscripts i and j refer to individual electrons and $r_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$. The first summation accounts for the kinetic energy of the electrons and their Coulomb interaction with the nucleus, while the second accounts for the electron-electron repulsion.

It is not possible to find an exact solution to the Schrödinger equation with a Hamiltonian of the form given by eqn 4.1, because the electron-electron repulsion term depends on the co-ordinates of two of the electrons, and so we cannot separate the wave function into a product of single-particle states. Furthermore, the electron-electron repulsion term is comparable in magnitude to the first summation, making it impossible to use perturbation theory either. The description of multi-electron atoms therefore usually starts with the **central field approximation** in which we re-write the Hamiltonian of eqn 4.1 in the form:[1]

$$\hat{H} = \sum_{i=1}^{N} \left(-\frac{\hbar^2}{2m} \boldsymbol{\nabla}_i^2 + V_{\text{central}}(r_i) \right) + V_{\text{residual}} \,, \tag{4.2}$$

where V_{central} is the **central field** and V_{residual} is the **residual electrostatic interaction**.

The central field approximation works in the limit where

$$\left| \sum_{i=1}^{N} V_{\text{central}}(r_i) \right| \gg |V_{\text{residual}}| \,. \tag{4.3}$$

In this case, we can treat V_{residual} as a perturbation, and worry about it later. We then have to solve a Schrödinger equation in the form:

$$\left[\sum_{i=1}^{N} \left(-\frac{\hbar^2}{2m} \boldsymbol{\nabla}_i^2 + V_{\text{central}}(r_i) \right) \right] \Psi = E\,\Psi \,. \tag{4.4}$$

[1] A field is described as "central" if the potential energy has spherical symmetry about the origin, so that $V(\boldsymbol{r})$ only depends on r. The fact that V does not depend on θ or ϕ means that the force is parallel to $\boldsymbol{r}$, i.e. it points *centrally* towards or away from the origin.

This is not as bad as it looks. By writing[2]

$$\Psi = \psi_1(\boldsymbol{r}_1)\,\psi_2(\boldsymbol{r}_2)\cdots\psi_N(\boldsymbol{r}_N)\,, \tag{4.5}$$

we end up with N separate Schrödinger equations of the form:

$$\left(-\frac{\hbar^2}{2m}\boldsymbol{\nabla}_i^2 + V_{\rm central}(r_i)\right)\psi_i(\boldsymbol{r}_i) = E_i\,\psi_i(\boldsymbol{r}_i)\,, \tag{4.6}$$

with

$$E = E_1 + E_2\cdots E_N\,. \tag{4.7}$$

This is much more tractable. We might need a computer to solve any one of the single particle wave Schrödinger equations of the type given in eqn 4.6, but at least it is possible in principle. Furthermore, the fact that the potentials that appear in eqn 4.6 only depend on the radial co-ordinate r_i (i.e. no dependence on the angles θ_i and ϕ_i) means that every electron is in a well-defined orbital angular momentum state,[3] and that the separation of variables discussed in Section 2.2 is valid. In analogy with eqn 2.25, we can then write:

$$\psi_i(\boldsymbol{r}_i) \equiv \psi(r_i,\theta_i,\phi_i) = R_i(r_i)\,Y_i(\theta_i,\phi_i)\,. \tag{4.8}$$

By proceeding exactly as in Section 2.2, we end up with two equations, namely:

$$\hat{\boldsymbol{L}}_i^{\,2} Y_{l_i m_i}(\theta_i,\phi_i) = \hbar^2 l_i(l_i+1) Y_{l_i m_i}(\theta_i,\phi_i)\,, \tag{4.9}$$

and

$$\left(-\frac{\hbar^2}{2m}\frac{1}{r_i^2}\frac{\mathrm{d}}{\mathrm{d}r_i}\left(r_i^2\frac{\mathrm{d}}{\mathrm{d}r_i}\right) + \frac{\hbar^2 l_i(l_i+1)}{2mr_i^2} + V_{\rm central}(r_i)\right)R_i(r_i) = E_i\,R_i(r_i)\,. \tag{4.10}$$

The first tells us that the angular part of the wave functions will be given by the spherical harmonic functions described in Section 2.2.3, while the second one allows us to work out the energy and radial wave function for a given form of $V_{\rm central}(r_i)$ and value of l_i. Each electron will therefore have four quantum numbers:

- l and m_l: these drop out of the angular equation for each electron, namely eqn 4.9.
- n: this arises from solving eqn 4.10 with the appropriate form of $V_{\rm central}(r)$ for a given value of l. n and l together determine the radial wave function $R_{nl}(r)$ (which cannot be expected to be the same as the hydrogenic ones given in Table 2.3) and the energy of the electron.
- m_s: spin has not entered the argument. Each electron can therefore either have spin up ($m_s = +1/2$) or down ($m_s = -1/2$), as usual. We do not need to specify the spin quantum number s because it is always equal to $1/2$.

The state of the many-electron atom is then found by working out the wave functions of the individual electrons and finding the total energy of the atom according to eqn 4.7, subject to the constraints imposed by the Pauli exclusion principle. This provides a useful working model that will be explored in detail below.

In the following sections we shall consider the experimental evidence for the shell model which proves that the central approximation is a good one. The reason it works is based on the nature of the shells. An individual electron experiences an electrostatic potential due to the Coulomb repulsion from all the other electrons in the atom. Nearly all of the electrons in a many-electron atom are in closed sub-shells, which have spherically-symmetric charge clouds. The off-radial forces from electrons in these closed shells cancel because of the spherical symmetry. Furthermore, the off-radial forces from electrons in unfilled shells are usually relatively small compared to the radial ones. We therefore expect the approximation given in eqn 4.3 to be valid for most atoms.

4.2 The shell model and the periodic table

We summarize here what we know so far about atomic states.

[2] The fact that electrons are indistinguishable particles means that we cannot distinguish physically between the case with electron 1 in state 1, electron 2 in state 2, ..., and the case with electron 2 in state 1, electron 1 in state 2, ..., etc. We should therefore really write down a *linear combination* of all such possibilities. We shall reconsider this point when considering the helium atom in Chapter 6.

[3] As noted in Section 2.2.3, the torque on the electron is zero if the force points centrally towards the nucleus. This means that the orbital angular momentum is constant.

Quantum number	symbol	Value
principal	n	any integer > 0
orbital	l	integer up to $(n-1)$
magnetic	m_l	integer from $-l$ to $+l$
spin	m_s	$\pm 1/2$

Table 4.1: Quantum numbers for electrons in atoms.

1. The electronic states are specified by four quantum numbers: n, l, m_l and m_s. The values that these quantum numbers can take are summarized in Table 4.1. In spectroscopic notation, electrons with $l = 0, 1, 2, 3, \ldots$ are called s, p, d, f, $\ldots$ electrons.

2. The gross energy of the electron is determined by n and l, except in hydrogenic atoms, where the gross structure depends only on n.

3. In the absence of fine structure and external magnetic fields, all the states with the same values of n and l are degenerate. Each (n, l) term of the gross structure therefore contains $2(2l+1)$ degenerate levels.

4. Electrons are indistinguishable, spin 1/2 particles and are therefore **fermions**. This means that they obey the **Pauli exclusion principle**, so that only one electron can occupy a particular quantum state.[4]

In the **shell model** of multi-electron atoms, we forget about fine structure and external magnetic fields, and just concentrate on the gross structure.[5] The energy levels are ordered according to the quantum numbers n and l, with big jumps in energy each time we move to the next set of quantum numbers. The degenerate states with the same values of n and l are called shells. As we add electrons to the atom, the Pauli exclusion principle dictates that the electrons fill up the lowest available shell until it is full, and then go on to the next one. The filling up of the shells in order of increasing energy in multi-electron atoms is sometimes called the **Aufbau principle**,[6] and is the basis of the **periodic table** of elements. The shells are listed in order of increasing energy in Table 4.2.

Inspection of Table 4.2 shows us that the energy of the shells always increases with n and l. We build up multi-electron atoms by adding electrons one by one, putting each electron into the lowest energy shell that has unfilled states. In general, this will be the one with the lowest n, but there are exceptions to this rule. For example, the 19^{th} electron goes into 4s shell rather than the 3d shell. Similarly, the 37^{th} electron goes into 5s shell rather than the 4d shell. This happens because the energy of the shell with a large l value may be higher than that of another shell with a larger value of n but smaller value of l.

The periodic table of elements is built up by adding electrons into the shells as the atomic number increases. This allows us to determine the **electronic configuration** of the elements, that is, the quantum numbers of the electrons in the atom. The configurations of the first 11 elements are listed in Table 4.3. The superscript attached to the shell tells us how many electrons are in that shell. The process of filling the shells follows the pattern indicated in Table 4.2. The nl sub-shells are filled *diagonally* when laid out in rows determined by the principal quantum number n, as shown in Fig. 4.1.[7]

4.3 Justification of the shell model

The theoretical justification for the shell model relies on the concept of **screening**. The idea is that the electrons in the inner shells screen the outer electrons from the potential of the nucleus. To see how this works we take sodium as an example.

[4] We shall discuss how the Pauli exclusion principle gives rise to exchange energy shifts in Chapter 6.

[5] This approximation is justified by the fact that the fine structure and magnetic field splittings are smaller than the gross structure energies by a factor of about $Z^2\alpha^2 = (Z^2/137)^2 \sim 10^{-4}\, Z^2$. Note, however, that the fine structure energy can get to be quite significant for large Z.

[6] The German word *Aufbau* means "building up".

[7] There are some exceptions to the general rules. For example, copper (Cu) with $Z = 29$ has a configuration of $\cdots 4s^1\ 3d^{10}$ instead of $\cdots 4s^2\ 3d^9$. This happens because filled shells are particularly stable. It is therefore energetically advantageous to promote the second 4s electron into the 3d shell to give the very stable $3d^{10}$ configuration. The energy difference between the two configurations is not particularly large, which explains why copper sometimes behaves as though it is monovalent, and sometimes divalent.

Shell	n	l	m_l	m_s	$\mathsf{N}_{\text{shell}}$	$\mathsf{N}_{\text{accum}}$
1s	1	0	0	$\pm 1/2$	2	2
2s	2	0	0	$\pm 1/2$	2	4
2p	2	1	$-1, 0, +1$	$\pm 1/2$	6	10
3s	3	0	0	$\pm 1/2$	2	12
3p	3	1	$-1, 0, +1$	$\pm 1/2$	6	18
4s	4	0	0	$\pm 1/2$	2	20
3d	3	2	$-2, -1, 0, +1, +2$	$\pm 1/2$	10	30
4p	4	1	$-1, 0, +1$	$\pm 1/2$	6	36
5s	5	0	0	$\pm 1/2$	2	38
4d	4	2	$-2, -1, 0, +1, +2$	$\pm 1/2$	10	48
5p	5	1	$-1, 0, +1$	$\pm 1/2$	6	54
6s	6	0	0	$\pm 1/2$	2	56
4f	4	3	$-3, -2, -1, 0, +1, +2, +3$	$\pm 1/2$	14	70
5d	5	2	$-2, -1, 0, +1, +2$	$\pm 1/2$	10	80
6p	6	1	$-1, 0, +1$	$\pm 1/2$	6	86
7s	7	0	0	$\pm 1/2$	2	88

Table 4.2: Atomic shells, listed in order of increasing energy. $\mathsf{N}_{\text{shell}}$ is equal to $2(2l+1)$ and is the number of electrons that can fit into the shell due to the degeneracy of the m_l and m_s levels. The last column gives the accumulated number of electrons that can be held by the atom once the particular shell and all the lower ones have been filled.

Element	Atomic number	Electronic configuration
H	1	$1s^1$
He	2	$1s^2$
Li	3	$1s^2\ 2s^1$
Be	4	$1s^2\ 2s^2$
B	5	$1s^2\ 2s^2\ 2p^1$
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	$1s^2\ 2s^2\ 2p^6\ 3s^1$

Table 4.3: The electronic configuration of the first 11 elements of the periodic table.

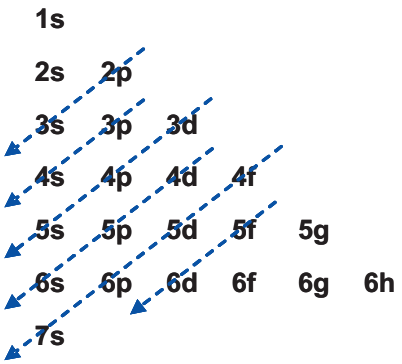

Figure 4.1: Atomic shells are filled in *diagonal* order when listed in rows according to the principal quantum number n.

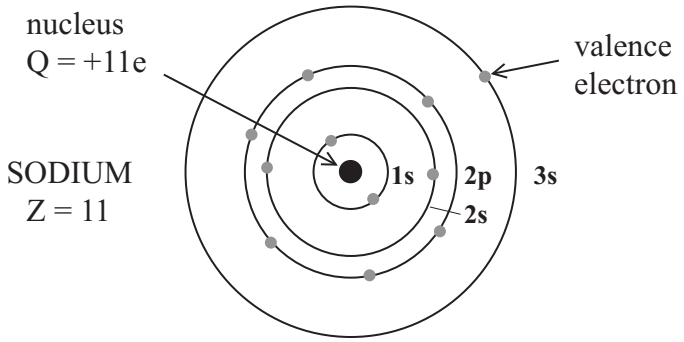

Figure 4.2: The electronic configuration of the sodium atom according to the shell model.

Shell	n	Z_{eff}	radius (Å)	Energy (eV)
1s	1	11	0.05	−1650
2s, 2p	2	9	0.24	−275
3s	3	1	4.8	−1.5

Table 4.4: Radii and energies of the principal atomic shells of sodium according to the Bohr model. The unit of 1 Ångstrom (Å) = 10^{-10} m.

Sodium has an atomic number of 11, and therefore has a nucleus with a charge of $+11e$ with 11 electrons orbiting around it. The picture of the atom based on the shell model is shown in Fig. 4.2. The radii and energies of the electrons in their shells are estimated using the Bohr formulæ:

$$r_n = \frac{n^2}{Z} a_{\text{H}}\,, \tag{4.11}$$

$$E_n = -\left(\frac{Z}{n}\right)^2 R_{\text{H}}\,, \tag{4.12}$$

where $a_{\text{H}} = 5.29 \times 10^{-11}$ m is the Bohr radius of hydrogen, $R_{\text{H}} = 13.6$ eV is the Rydberg constant and Z is the atomic number.

The first two electrons go into the $n = 1$ shell. These electrons see the full nuclear charge of $+11e$. With $n = 1$ and $Z = 11$, we find $r_1 = 1^2/11 \times a_{\text{H}} = 0.05\,\text{Å}$ and $E_1 = -11^2 R_{\text{H}} = -1650$ eV. The next eight electrons go into the $n = 2$ shell. These are presumed to orbit outside the $n = 1$ shell. The two inner electrons partly screen the nuclear charge, and the $n = 2$ electrons see an effective charge $Z_{\text{eff}} = +9e$. The radius is therefore $r_2 = (2^2/9) \times a_{\text{H}} = 0.24\,\text{Å}$ and the energy is $E_2 = -(9/2)^2 R_{\text{H}} = -275$ eV. Finally, the outermost electron in the $n = 3$ shell orbits outside the filled $n = 1$ and $n = 2$ shells, and therefore sees $Z_{\text{eff}} = 1$. With $Z = 1$ and $n = 3$ we find $r_3 = 4.8\,\text{Å}$ and $E_3 = -1.5$ eV. These values are summarized in Table 4.4. Note the large jump in energy and radius in moving from one shell to the next.

The treatment of the screening discussed in the previous paragraph is clearly over-simplified because it is based on Bohr-type orbits and does not treat the electron-electron repulsion properly. In Section 4.5 we shall see how we might improve on it. One point to realize, however, is that the model is reasonably self-consistent: by assuming that the inner shells screen the outer ones, we find that the orbital radius increases in each subsequent shell, which corroborates our original assumption. This is why the model works so well.

4.4 Experimental evidence for the shell model

There is a wealth of experimental evidence to confirm that the shell model is a good one. The main points are discussed briefly here.

The periodic table of elements

The periodic table of elements follows from the electronic configuration of the elements, which is derived from the shell structure of atoms. The periodic table underpins the chemical activity of the elements.

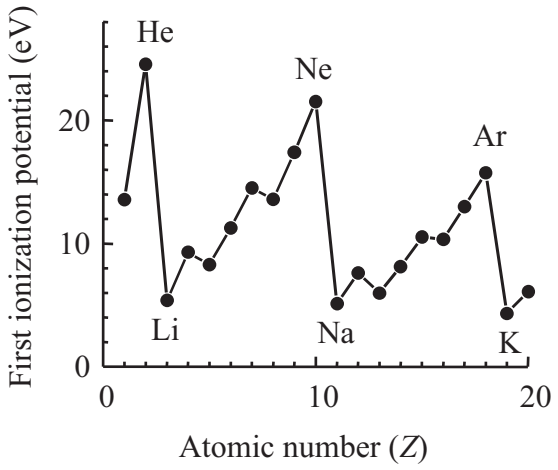

Figure 4.3: First ionization potentials of the elements up to calcium. The noble gas elements (He, Ne, Ar) have highly stable fully filled shells with large ionization potentials. The alkali metals (Li, Na, K) have one weakly-bound valence electron outside fully-filled shells.

It can thus be argued that the whole subject of chemistry can be regarded as experimental proof of the shell structure of atoms.

Ionization potentials and atomic radii

The ionization potentials of the noble gas elements are the highest within a particular period of the atomic table, while those of the alkali metals are the lowest. This can be seen by looking at the data in Fig. 4.3. The ionization potential gradually increases as the atomic number increases until the shell is filled, and then it drops abruptly. This shows that the filled shells are very stable, and that the valence electrons go in larger, less tightly-bound orbits. The results correlate with the chemical activity of the elements. The noble gases require large amounts of energy to liberate their outermost electrons, and they are therefore chemically inert. The alkali metals, on the other hand, need much less energy, and are therefore highly reactive.

It is also found that the average atomic radius determined by X-ray crystallography on closely packed crystals is largest for the alkali metals. This is further evidence that we have weakly-bound valence electrons outside strongly-bound, small-radius, inner shells.

X-ray line spectra

Measurements of X-ray line spectra allow the energies of the inner shells to be determined directly. The experimental arrangement for observing an X-ray emission spectrum is shown in Fig. 4.4(a). Electrons are accelerated across a potential drop of several kV and then impact on a target. This ejects **core electrons** from the inner shells of the target, as shown in Fig. 4.4(b). X-ray photons are emitted as the higher energy electrons drop down to fill the empty level (or **hole**) in the lower shell.

Each target emits a series of characteristic lines. The series generated when a K-shell ($n = 1$) electron has been ejected is called the K-series. Similarly, the L- and M-series correspond to ejection of L-shell ($n = 2$) or M-shell ($n = 3$) core electrons respectively. This old spectroscopic notation dates back to the early work on X-ray spectra.

Figure 4.5(a) shows a typical X-ray emission spectrum. The spectrum consists of a series of sharp lines on top of a continuous spectrum. The groups of sharp lines are generated by radiative transitions following the ejection of an inner shell electron as indicated in Fig. 4.4(b). The group of lines around 0.2 Å originate from K-shell transitions, while the three groups of lines between 1.0 Å and 1.6 Å arise from L-shell transitions. A particular set of lines is only observed if the tube voltage is high enough to eject the relevant electron. Hence new groups of lines appear as the voltage is increased, as the higher energy electron beam ejects ever deeper inner shell electrons. At a given voltage, several groups of lines are observed as the hole in the initial shell moves up through the higher shells. For example, L-shell lines are observed at the same time as K-shell lines after the electron in the L-shell drops to the hole in the K-shell, thus leaving a hole in the L-shell, and so on.

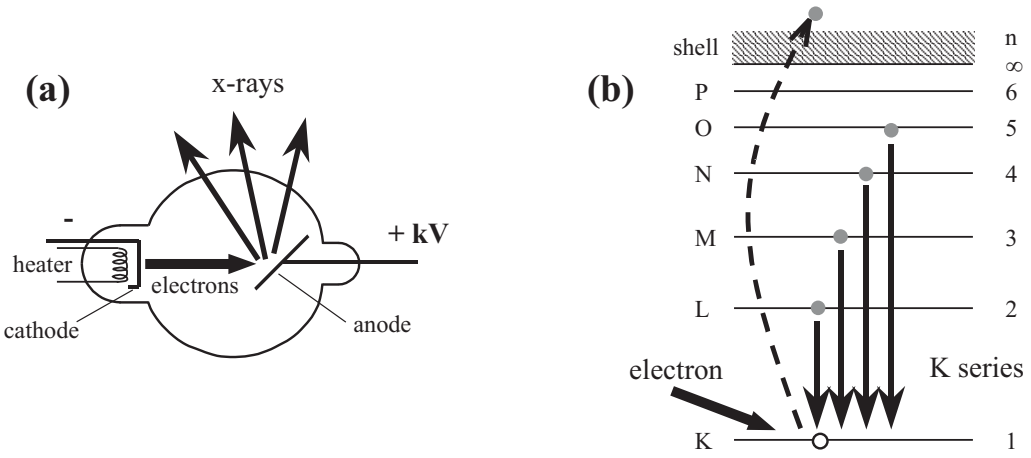

Figure 4.4: (a) A typical X-ray tube. Electrons are accelerated with a voltage of several kV and impact on a target, causing it to emit X-rays. (b) Transitions occurring in the K-series emission lines. An electron from the discharge tube ejects one of the K-shell electrons of the target, leaving an empty level in the K shell. X-ray photons are emitted as electrons from the higher shells drop down to fill the hole in the K-shell.

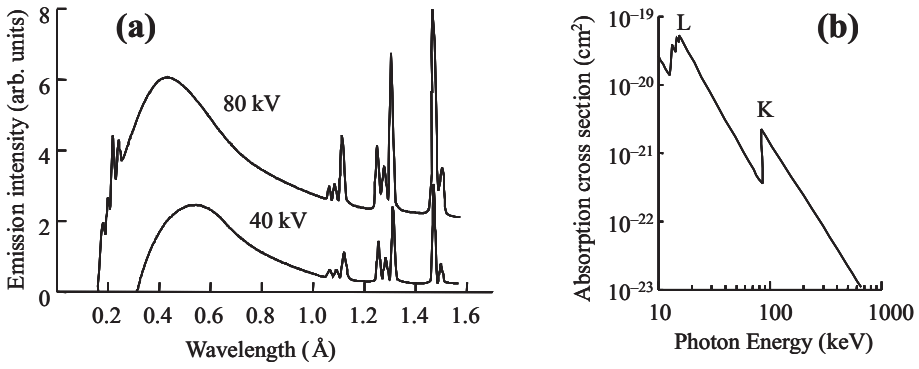

Figure 4.5: (a) X-ray emission spectra for tungsten at two different electron voltages. The sharp lines are caused by radiative transitions after the electron beam ejects an inner shell electron, as indicated in Fig. 4.4(b). The continuum is caused by bremsstrahlung, which has a short wavelength limit equal to hc/eV at voltage V. (b) X-ray absorption cross-section spectrum for lead.

The continuous spectrum is caused by **bremsstrahlung**.[8] Bremsstrahlung occurs when the electron is scattered by the atoms without ejecting a core electron from the target. The acceleration of the electron associated with its change of direction causes it to radiate. Conservation of energy demands that the frequency of the radiation must cut off when $h\nu = eV$, V being the voltage across the tube. This means that the minimum wavelength is equal to hc/eV. The reduction of the short wavelength limit of the bremsstrahlung with increasing voltage is apparent in the data shown in Fig. 4.5(a).

The energy of an electron in an inner shell with principal quantum number n is given by:

$$E_n = -\frac{{Z_n^{\rm eff}}^2}{n^2} R_{\rm H}\,, \tag{4.13}$$

where $Z_n^{\rm eff}$ is the effective nuclear charge, and $R_{\rm H} = 13.6\,{\rm eV}$. The difference between Z (the atomic number of the target) and $Z_n^{\rm eff}$ is caused by the screening effect of the other electrons. The energy of the optical transition from $n \to n'$ is thus given by:

$$h\nu = |E_{n'} - E_n| = \left| \frac{{Z_n^{\rm eff}}^2}{n^2} - \frac{{Z_{n'}^{\rm eff}}^2}{{n'}^2} \right| R_{\rm H}\,. \tag{4.14}$$

In practice, the wavelengths of the various series of emission lines are found to obey **Moseley's law**, where we make the approximation $Z_n^{\rm eff} = Z_{n'}^{\rm eff}$ and write both as $(Z - \sigma_n)$. For example, the K-shell lines are given by:[9]

$$\frac{hc}{\lambda} \approx (Z - \sigma_K)^2 R_H \left(\frac{1}{1^2} - \frac{1}{n^2} \right)\,, \tag{4.15}$$

where $n > 1$ and $\sigma_K \sim 3$. Similarly, the L-shell spectra obey:

$$\frac{hc}{\lambda} \approx (Z - \sigma_L)^2 R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right)\,, \tag{4.16}$$

where $n > 2$, and $\sigma_L \sim 10$. We can see that these are just the expected wavelengths predicted by the Bohr model, except that we have an effective charge of $(Z - \sigma_n)$ instead of Z. The phenomenological **screening parameter** σ_n that appears here accounts for the screening of the nucleus by the other electrons and varies from shell to shell.

X-ray absorption spectra show a complementary frequency dependence to the emission spectra. Figure 4.5(b) shows a typical X-ray absorption spectrum. A sharp increase in the absorption cross section[10] occurs whenever the photon energy crosses the threshold to eject an electron out of an inner shell to empty states above the highest occupied shell. This sharp increase in the absorption is called the **absorption edge**. The final state for the electron after the absorption transition could either be one of the excited states of the valence electrons or in the continuum above the ionization limit. The binding energy of the valence electrons is negligible on the scale of X-ray energies, and so we can effectively put $E_{n'} = 0$ in eqn 4.14 and hence obtain the energy of the absorption edge as:

$$h\nu^{\rm edge} = \frac{{Z_n^{\rm eff}}^2}{n^2} R_{\rm H} \equiv \frac{(Z - \sigma_n)^2}{n^2} R_{\rm H}\,. \tag{4.17}$$

The absorption probability decreases as the electron gets promoted further into the continuum. Hence we see a peak in the absorption at $h\nu^{\rm edge}$ and a decrease thereafter. The K and L shell absorption edges are clearly visible at 88 keV and 15 keV respectively in Fig. 4.5(b).

Close inspection of Fig. 4.5(b) reveals that there is some sub-structure in the L-shell absorption edge, but not in the K-shell edge. This is a consequence of the fact that the L-shell has two sub-shells corresponding to different values of the angular momentum quantum number l, namely the 2s and 2p sub-shells. These have slightly different screening parameters, and hence slightly different energies, on account of the different shape of their radial wave functions. The K shell, by contrast, can only have $l = 0$, and thus consists a unique state, namely the 1s level. The situation is further complicated by the

[8] German: *brems* = braking (i.e. deceleration) + *strahlung* = radiation.

[9] There is no real scientific justification for the approximation $Z_n^{\rm eff} = Z_{n'}^{\rm eff}$ in Moseley's law. The law is an empirical one and reflects the fact that the transition wavelength is mainly dominated by the energy of the lower shell. Note also that close inspection of the line spectra reveals sub-structure due to the relatively small energy differences between the l states for a particular value of n, and further smaller splittings due to spin-orbit coupling.

[10] Absorption coefficients are often expressed as "cross sections". The cross section is equal to the effective area of the beam that is blocked out by the absorption of an individual atom. If there are N atoms per unit volume, and the cross section is equal to $\sigma_{\rm abs}$, the absorption coefficient in ${\rm m}^{-1}$ is equal to $N\sigma_{\rm abs}$.

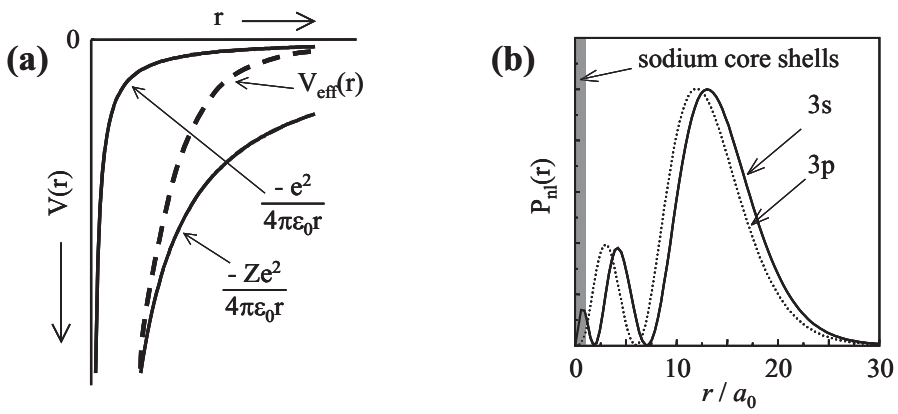

Figure 4.6: (a) Typical effective potential $V_{\rm eff}(r)$ for the valence electrons of an atom with atomic number Z. (b) Radial probability densities for hydrogenic 3s and 3p wave functions. a_0 is the Bohr radius (0.529 Å). The shaded region near $r = 0$ represents the inner core shells for the case of sodium with $Z = 11$.

fact that the 2p sub-shell is split by the spin-orbit effect that will be discussed in Chapter 7, which means that the L-shell actually consists of *three* sub-shells.[11] This explains why the L-edge around 20 keV has three sub-edges in Fig. 4.5(b), and why there are three groups of lines in the emission spectrum shown in Fig. 4.5(a) for the L-shell transitions from 1.0–1.6Å.

Detailed lists of X-ray transition energies may be found on the NIST atomic data base. [12] Another useful reference is maintained by the National Physical Laboratory (NPL) in the United Kingdom.

4.5 Effective potentials, screening, and alkali metals

The electrons in a multi-electron atom arrange themselves with the smallest number of electrons in unfilled shells outside inner filled shells. These outermost electrons are called the **valence electrons** of the atom. They are responsible for the chemical activity of the particular elements.

In order to work out the energy levels of the valence electrons, we need to solve the N-electron Schrödinger equation given in eqn 4.1. Within the central-field approximation, each valence electron satisfies a Schrödinger equation of the type given in eqn 4.6, which can be written in the form:

$$\left(-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V^l_{\rm eff}(r)\right)\psi = E\,\psi\,. \tag{4.18}$$

The Coulomb repulsion from the core electrons is lumped into the effective potential $V^l_{\rm eff}(r)$. This is only an approximation to the real behaviour, but it can be reasonably good, depending on how well we work out $V^l_{\rm eff}(r)$. Note that the effective potential depends on l. This arises from the term in l that appears in eqn 4.10 and has important consequences, as we shall see below.

The overall dependence of $V_{\rm eff}(r)$ with r must look something like Fig. 4.6. At very large values of r, the outermost valence electron will be well outside any filled shells, and will thus only see an attractive potential equivalent to a charge of $+e$. On the other hand, if r is very small, the electron will see the full nuclear charge of $+Ze$. The potential at intermediate values of r must lie somewhere between these two limits: hence the generic form of $V_{\rm eff}(r)$ shown in Fig. 4.6. The task of calculating $V^l_{\rm eff}(r)$ keeps theoretical atomic physicists busy. Two common approximation techniques used to perform the calculations are called the **Hartree** and **Thomas-Fermi** methods.

As a specific example, we consider the **alkali metals** such as lithium, sodium and potassium, which come from group I of the periodic table. They have one valence electron outside filled inner shells, as indicated in Table 4.5. They are therefore approximately one-electron systems, and can be understood by introducing a phenomenological number called the **quantum defect** to describe the energies. To see how this works, we consider the sodium atom.

[11] The spin-orbit effect is zero for s-states such as the 1s and 2s sub-shells, because they have $l = 0$. See eqn 7.35.

[12] See http://www.nist.gov/pml/data/xray_gammaray.cfm. The National Physical Laboratory (NPL) maintains an online, updated version of the classic data reference book by G.W.C. Kaye and T.H. Laby, which includes detailed X-ray data. See http://www.kayelaby.npl.co.uk.

Element	Z	Electronic configuration
Lithium	3	$1s^2 2s^1$
Sodium	11	[Ne] $3s^1$
Potassium	19	[Ar] $4s^1$
Rubidium	37	[Kr] $5s^1$
Cesium	55	[Xe] $6s^1$

Table 4.5: Alkali metals. The symbol [...] indicates that the inner shells are filled according to the electronic configuration of the noble gas element identified in the bracket.

l	$n=3$	$n=4$	$n=5$	$n=6$
0	1.373	1.357	1.352	1.349
1	0.883	0.867	0.862	0.859
2	0.010	0.011	0.013	0.011
3	–	0.000	-0.001	-0.008

Table 4.6: Values of the quantum defect $\delta(l)$ for sodium against n and l.

The shell model picture of sodium is shown in Fig. 4.2. The optical spectra are determined by excitations of the outermost 3s electron. The energy of each (n, l) term of the valence electron is given by:

$$E_{nl} = -\frac{R_H}{[n-\delta(l)]^2}\,, \tag{4.19}$$

where $n \geq 3$ and $\delta(l)$ is the quantum defect. The quantum defect allows for the penetration of the inner shells by the valence electron.

The dependence of the quantum defect on l can be understood with reference to Fig. 4.6(b). This shows the radial probability densities $P_{nl}(r) = r^2|R(r)|^2$ for the 3s and 3p orbitals of a hydrogenic atom with $Z = 1$, which might be expected to be a reasonable approximation for the single valence electron of sodium. The shaded region near $r = 0$ represents the inner $n = 1$ and $n = 2$ shells with radii of $\sim 0.09a_0$ and $\sim 0.44a_0$ respectively. (See Section 4.3.) We see that both the 3s and 3p orbitals penetrate the inner shells, and that this penetration is much greater for the 3s electron. The electron will therefore see a larger effective nuclear charge for part of its orbit, and this will have the effect of reducing the energies. The energy reduction is largest for the 3s electron due to its larger core penetration.

The quantum defect $\delta(l)$ was introduced empirically to account for the optical spectra. In principle it should depend on both n and l, but it was found experimentally to depend mainly on l. This can be seen from the values of the quantum defect for sodium tabulated in Table 4.6. The corresponding energy spectrum is shown schematically in Fig. 4.7. Note that $\delta(l)$ is very small for $l \geq 2$.

We can use the quantum defect to calculate the wavelengths of the emission lines. The D lines correspond to the 3p $\rightarrow$ 3s transition.[13] By using the values of δ given in Table 4.6, we can find the photon energy, and hence the photon wavelength, from the energy difference between the two levels:

$$h\nu = \frac{hc}{\lambda} = E_{3p} - E_{3s}\,. \tag{4.20}$$

We therefore find that the wavelength λ is given by:

$$\begin{aligned}\frac{1}{\lambda} &= \frac{R_H}{hc}\left(\frac{1}{[3-\delta(3\mathrm{s})]^2} - \frac{1}{[3-\delta(3\mathrm{p})]^2}\right) \\ &= (1.10\times 10^5\,\mathrm{cm}^{-1}) \times \left(\frac{1}{1.627^2} - \frac{1}{2.117^2}\right)\,.\end{aligned}$$

The wave number $\overline{\nu} \equiv 1/\lambda$ of the transition is thus $1.70 \times 10^4\,\mathrm{cm}^{-1}$, and so λ is equal to 590nm. This corresponds to the yellow-orange part of the spectrum, as we would expect for the D-lines of sodium.

[13] The labelling of the 3p $\rightarrow$ 3s transition as a "D-line" dates back to Fraunhofer's catalogue of spectral lines. The term is now applied to the first transition from the ground state of any alkali atom, i.e. the ns $\leftrightarrow$ np transition, where ns is the ground state of the alkali.

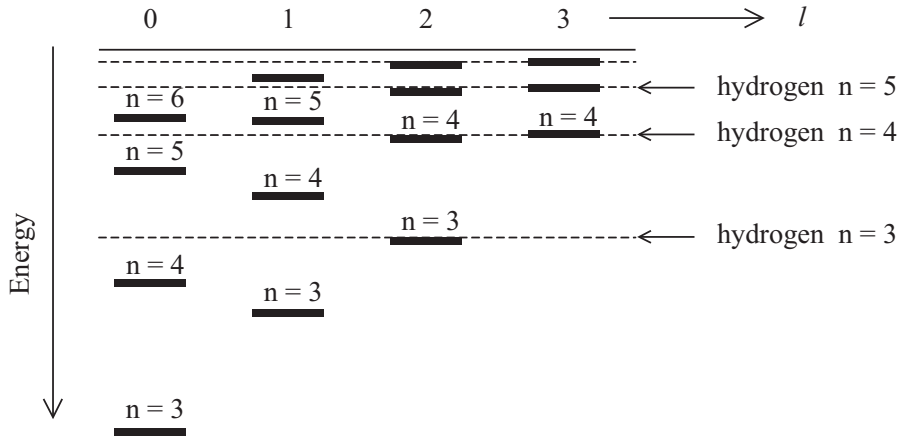

Figure 4.7: Schematic energy level diagram for sodium, showing the ordering of the energy levels.

Reading

Bransden and Joachain, *Physics of Atoms and Molecules*, §8.1, 8.2, 9.4, 9.7
Demtröder, *Atoms, Molecules and Photons*, §6.2–6.4, §7.5
Haken and Wolf, *The physics of atoms and quanta*, chapters 11, 18 & 19.
Hertel and Schulz, *Atoms, Molecules and Optical Physics, 1*, §3.1 – 3.2, 10.5
Phillips, A.C., *Introduction to Quantum Mechanics*, chapter 11.
Eisberg and Resnick, *Quantum Physics*, chapters 9 & 10.
Beisser, *Concepts of Modern Physics*, chapter 7.

Chapter 5

Angular momentum

The treatment of angular momentum is very important for understanding the properties of atoms. It is now time to explore these effects in detail, and to see how this leads to the classification of the quantized states of atoms by their angular momentum.

5.1 Conservation of angular momentum

In the Sections that follow, we are going to consider several different types of angular momentum, and the ways in which they are coupled together. Before going into the details, it is useful to stress one very important point related to conservation of angular momentum. In an isolated atom, there are many forces (and hence torques) acting inside the atom. These *internal* forces cannot change the *total* angular momentum of the atom, since conservation of angular momentum demands that the angular momentum of the atom as a whole must be conserved in the absence of any *external* torques. The total angular momentum of the atom is normally determined by its electrons. The total electronic angular momentum is written $\boldsymbol{J}$, and is specified by the quantum number J. The principle of conservation of angular momentum therefore requires that isolated atoms always have well-defined J states.[1] It is this J value that determines, for example, the magnetic dipole moment of the atom.

The principle of conservation of angular momentum does not apply, of course, when external perturbations are applied. The most obvious example is the perturbation caused by the emission or absorption of a photon. In this case the angular momentum of the atom must change because the photon itself carries angular momentum, and the angular momentum of the whole system (atom + photon) has to be conserved. The change in J is then governed by selection rules, as discussed, for example, in Section 5.8. Another obvious example is the effect of a strong external DC magnetic field. In this case it is possible for the magnetic field to produce states where the component of angular momentum along the direction of the field is well-defined, but not the total angular momentum. (See the discussion of the Paschen-Back effect in Section 8.1.3.)

5.2 Types of angular momentum

The electrons in atoms possess two different types of angular momentum, namely orbital and spin angular momentum.These are discussed separately below.

5.2.1 Orbital angular momentum

The electrons in atoms orbit around the nucleus, and therefore possess orbital angular momentum. In classical mechanics, we define the orbital angular momentum of a particle by:

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}\,, \tag{5.1}$$

[1] This statement about J has to be qualified somewhat when we add in the effects of the nucleus. The angular momentum of an atom is the resultant of the electronic angular momentum and the nuclear spin. The total angular momentum of an isolated atom has to be conserved, but the electrons can exchange angular momentum with the nucleus through **hyperfine** interactions. (See Section 7.7.2.) These interactions are very weak, and can usually be neglected except when explicitly considering nuclear effects.

where $\boldsymbol{r}$ is the radial position, and $\boldsymbol{p}$ is the linear momentum. The components of $\boldsymbol{L}$ are given by

$$\begin{pmatrix} L_x \\ L_y \\ L_z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \times \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} yp_z - zp_y \\ zp_x - xp_z \\ xp_y - yp_x \end{pmatrix} . \tag{5.2}$$

In quantum mechanics we represent the linear momentum by differential operators of the type

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x} . \tag{5.3}$$

Therefore, the quantum mechanical operators for the Cartesian components of the orbital angular momentum are given by:

$$\hat{L}_x = -i\hbar \left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \right) \tag{5.4}$$

$$\hat{L}_y = -i\hbar \left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \right) \tag{5.5}$$

$$\hat{L}_z = -i\hbar \left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \right) . \tag{5.6}$$

Note that the "hat" symbol indicates that we are representing an operator and not just a number.

In classical mechanics, the magnitude of the angular momentum is given by:

$$L^2 = L_x^2 + L_y^2 + L_z^2 .$$

We therefore define the quantum mechanical operator for the magnitude of the angular momentum by:

$$\hat{\boldsymbol{L}}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 . \tag{5.7}$$

The operators like $\hat{L}_x^2$ that appear here should be understood in terms of repeated operations:

$$\begin{aligned} \hat{L}_x^2\psi &= -\hbar^2 \left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \right) \left(y\frac{\partial \psi}{\partial z} - z\frac{\partial \psi}{\partial y} \right) \\ &= -\hbar^2 \left(y^2\frac{\partial^2 \psi}{\partial z^2} - y\frac{\partial \psi}{\partial y} - z\frac{\partial \psi}{\partial z} - 2yz\frac{\partial^2 \psi}{\partial y \partial z} + z^2\frac{\partial^2 \psi}{\partial y^2} \right) . \end{aligned}$$

Note that we have already met the $\hat{\boldsymbol{L}}^2$ and $\hat{L}_z$ operators when we solved the Schrödinger equation for hydrogen in Section 2.2. (See eqns 2.27 and 2.33.) When considering hydrogen, the spherical symmetry of the atom made it convenient to work in spherical polar rather than Cartesian co-ordinates. The two approaches are, of course, completely equivalent, and the operators are physically identical, whether expressed in their spherical polar or Cartesian forms.

A key property of the orbital angular momentum operator is that its components do not commute with each other, but they do commute with $\hat{\boldsymbol{L}}^2$. We can summarise this by writing the commutators:[2]

$$\begin{aligned} [\hat{L}_x, \hat{L}_y] &\neq 0 , \\ [\hat{\boldsymbol{L}}^2, L_z] &= 0 . \end{aligned} \tag{5.8}$$

The non-commutation of the components can be proved as follows:

$$\begin{aligned} \hat{L}_x\hat{L}_y\psi &= (-i\hbar)^2 \left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \right) \left(z\frac{\partial \psi}{\partial x} - x\frac{\partial \psi}{\partial z} \right) , \\ &= -\hbar^2 \left(yz\frac{\partial^2 \psi}{\partial z \partial x} + y\frac{\partial \psi}{\partial x} - yx\frac{\partial^2 \psi}{\partial z^2} - z^2\frac{\partial^2 \psi}{\partial y \partial x} + zx\frac{\partial^2 \psi}{\partial y \partial z} \right) . \end{aligned}$$

On the other hand, we have:

$$\begin{aligned} \hat{L}_y\hat{L}_x\psi &= (-i\hbar)^2 \left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \right) \left(y\frac{\partial \psi}{\partial z} - z\frac{\partial \psi}{\partial y} \right) , \\ &= -\hbar^2 \left(zy\frac{\partial^2 \psi}{\partial x \partial z} - z^2\frac{\partial^2 \psi}{\partial x \partial y} - xy\frac{\partial^2 \psi}{\partial z^2} + xz\frac{\partial^2 \psi}{\partial z \partial y} + x\frac{\partial \psi}{\partial y} \right) . \end{aligned}$$

[2]The commutator of two quantum mechanical operators $\hat{A}$ and $\hat{B}$ is defined by: $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$. Hence $[\hat{L}_x, \hat{L}_y] = \hat{L}_x\hat{L}_y - \hat{L}_y\hat{L}_x$.

On recalling that $\partial^2\psi/\partial x\partial y = \partial^2\psi/\partial y\partial x$, we find:

$$\begin{aligned}\hat{L}_x\hat{L}_y\psi - \hat{L}_y\hat{L}_x\psi \equiv [\hat{L}_x, \hat{L}_y]\psi &= -\hbar^2\left(y\frac{\partial\psi}{\partial x} - x\frac{\partial\psi}{\partial y}\right), \\ &= i\hbar \times -i\hbar\left(x\frac{\partial\psi}{\partial y} - y\frac{\partial\psi}{\partial x}\right), \\ &= i\hbar\hat{L}_z\psi\,.\end{aligned}$$

We therefore conclude that:

$$[\hat{L}_x, \hat{L}_y] = i\hbar\hat{L}_z \tag{5.9}$$

The other commutators of the angular momentum operators, namely $[\hat{L}_y, \hat{L}_z]$ and $[\hat{L}_z, \hat{L}_x]$ are obtained by cyclic permutation of the indices in Eq. 5.9: $x \to y$, $y \to z$, $z \to x$.

The commutation of $\hat{\boldsymbol{L}}^2$ with $\hat{L}_z$ (i.e. $[\hat{\boldsymbol{L}}^2, \hat{L}_z] = 0$) can be proven by a number of ways. Here is one. We use the identity:[3]

$$[\hat{A}^2, \hat{B}] = \hat{A}[\hat{A}, \hat{B}] + [\hat{A}, \hat{B}]\hat{A} \tag{5.10}$$

together with eqn 5.7 and the cyclic permutations of eqn 5.9 to write

$$\begin{aligned}[\hat{\boldsymbol{L}}^2, \hat{L}_z] &= [\hat{L}_x^2, \hat{L}_z] + [\hat{L}_y^2, \hat{L}_z] + [\hat{L}_z^2, \hat{L}_z]\,, \\ &= [\hat{L}_x^2, \hat{L}_z] + [\hat{L}_y^2, \hat{L}_z] + 0\,, \\ &= \hat{L}_x[\hat{L}_x, \hat{L}_z] + [\hat{L}_x, \hat{L}_z]\hat{L}_x + \hat{L}_y[\hat{L}_y, \hat{L}_z] + [\hat{L}_y, \hat{L}_z]\hat{L}_y\,, \\ &= -i\hbar\hat{L}_x\hat{L}_y - i\hbar\hat{L}_y\hat{L}_x + i\hbar\hat{L}_y\hat{L}_x + i\hbar\hat{L}_x\hat{L}_y\,, \\ &= 0\,.\end{aligned}$$

It can be shown that the measurable quantities corresponding to two quantum mechanical operators that do not commute must obey an uncertainty principle. The general result for operators $\hat{A}$ and $\hat{B}$ is:

$$\Delta A^2\Delta B^2 \geq \frac{1}{4}\left|\langle[\hat{A}, \hat{B}]\rangle\right|^2\,. \tag{5.11}$$

The Heisenberg uncertainty principle $\Delta x\Delta p \geq \hbar/2$ is a well known example of this.[4] The non-commutation of the components of $\boldsymbol{L}$ thus implies that it is not possible to know the values of L_x, L_y, L_z simultaneously: we can only know *one* of them (usually L_z) at any time. Once L_z is known, we cannot know L_x and L_y as well. On the other hand, the fact that $\hat{L}_z$ commutes with $\hat{\boldsymbol{L}}^2$ (cf. eqn 5.8) means that we *can* know the length of the angular momentum vector and its z component simultaneously. In summary:

- We can know the length of the angular momentum vector L and one of its components.
- For mathematical convenience, we usually take the component we know to be L_z.
- We cannot know the values of all three components of the angular momentum simultaneously.

The eigenvalues of the angular momentum operators were discussed in Section 2.2.3. The orbital angular momentum is specified by two quantum numbers: l and m. The latter is sometimes given an extra subscript (i.e. m_l) to distinguish it from the spin quantum number m_s considered below. The magnitude of $\boldsymbol{l}$ is given by

$$|\boldsymbol{l}| = \sqrt{l(l+1)}\hbar\,, \tag{5.12}$$

and the component along the z axis by

$$l_z = m\hbar\,. \tag{5.13}$$

Note that we have switched to a lower case notation here because we are referring to a *single* electron. (See Section 5.7.) l can take positive integer values (including 0) and m can take values in integer steps

[3] This is proven as follows:

$$[\hat{A}^2, \hat{B}] \equiv \hat{A}\hat{A}\hat{B} - \hat{B}\hat{A}\hat{A} = \hat{A}\hat{A}\hat{B} - \hat{A}\hat{B}\hat{A} + \hat{A}\hat{B}\hat{A} - \hat{B}\hat{A}\hat{A} = \hat{A}(\hat{A}\hat{B} - \hat{B}\hat{A}) + (\hat{A}\hat{B} - \hat{B}\hat{A})\hat{A} \equiv \hat{A}[\hat{A}, \hat{B}] + [\hat{A}, \hat{B}]\hat{A}.$$

[4] The commutator of $\hat{x}$ and $\hat{p}$ is given by:

$$[\hat{x}, \hat{p}]\psi = (\hat{x}\hat{p} - \hat{p}\hat{x})\psi = -i\hbar\,x\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\right) + i\hbar\frac{\mathrm{d}(x\psi)}{\mathrm{d}x} = i\hbar\psi\,.$$

Hence $[\hat{x}, \hat{p}] = i\hbar$.

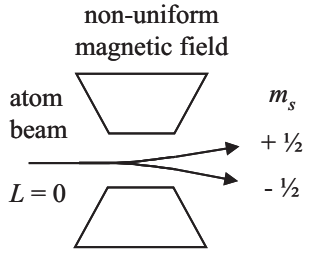

Figure 5.1: The Stern–Gerlach experiment. A beam of monovalent atoms with $L = 0$ (i.e. zero orbital angular momentum and hence zero orbital magnetic dipole moment) is deflected in two discrete ways by a non-uniform magnetic field. The force on the atoms arises from the interaction between the field and the magnetic moment due to the electron spin.

from $-l$ to $+l$. The number of m states for each l state is therefore equal to $(2l+1)$. These m states are degenerate in isolated atoms, but can be split by external perturbations (e.g. magnetic or electric fields.)

The quantisation of the angular momentum can be represented pictorially in the **vector model** , as shown previously in figure 2.3. In this model the angular momentum is represented as a vector of length $\sqrt{l(l+1)}\hbar$ angled so that its component along the z axis is equal to $m\hbar$. The x and y components of the angular momentum are not known.

In classical mechanics, the orbital angular momentum is conserved when the force $\boldsymbol{F}$ is radial: i.e. $\boldsymbol{F} \equiv F\hat{\boldsymbol{r}}$, where $\hat{\boldsymbol{r}}$ is a unit vector parallel to $\boldsymbol{r}$. This follows from the equation of motion:

$$\frac{\mathrm{d}\boldsymbol{l}}{\mathrm{d}t} = \boldsymbol{\Gamma} = \boldsymbol{r} \times \boldsymbol{F} = \boldsymbol{r} \times F\hat{\boldsymbol{r}} = 0\,, \tag{5.14}$$

where $\boldsymbol{\Gamma}$ is the torque. In the hydrogen atom, the Coulomb force on the electron acts towards the nucleus, and hence $\boldsymbol{l}$ is conserved. This is why the angular momentum ends up being quantized with well-defined constant values when we consider the quantum mechanics of the hydrogen atom. It is also the case that the individual electrons of many-electron atoms have well-defined l states. This follows because the central field approximation gives a very good description of the behaviour of many electron atoms (see Section 4.1), and the dominant resultant force on the electron is radial (i.e. central) in this limit.[5]

5.2.2 Spin angular momentum

A wealth of data derived from the optical, magnetic and chemical properties of atoms points to the fact that electrons possess an additional type of angular momentum called **spin**. The electron behaves as if it spins around its own internal axis, but this analogy should not be taken literally — the electron is, as far as we know, a point particle, and so cannot be spinning in any classical way. In fact, spin is a purely quantum effect with no classical explanation. Paul Dirac at Cambridge successfully accounted for electron spin when he produced the relativistic wave equation that bears his name in 1928.

The discovery of spin goes back to the Stern-Gerlach experiment, in which a beam of atoms is deflected by a non-uniform magnetic field. (See Fig. 5.1). The force on a magnetic dipole in a non-uniform magnetic field is given by:[6]

$$F_z = \mu_z \frac{\mathrm{d}B}{\mathrm{d}z}\,, \tag{5.15}$$

where $\mathrm{d}B/\mathrm{d}z$ is the field gradient, which is assumed to point along the z direction, and μ_z is the z-component of the magnetic dipole of the atom. In Chapter 7 we shall explore the origin of magnetic dipoles in detail. At this stage, all we need to know is that the magnetic dipole is directly proportional to the angular momentum of the atom. (See Section 7.1.)

[5] The inclusion of non-central forces via the residual electrostatic interaction leads to some mixing of the orbital angular momentum states. This can explain why transitions that are apparently forbidden by selection rules can sometimes be observed, albeit with low transition probabilities.

[6] Note that we need a *non-uniform* magnetic field to deflect a magnetic dipole. A *uniform* magnetic field merely exerts a torque, not a force. We can understand this by analogy with electrostatics. Electric monopoles (i.e. free charges) can be moved by applying electric fields, but an electric dipole experiences no net force in a uniform electric field because the forces on the positive and negative charges cancel. If we wish to apply a force to an electric dipole, we therefore need to apply a non-uniform electric field, so that the forces on the two charges are different. Magnetic monopoles do not exist (as far as we know), and so all atomic magnets are dipoles. Hence we must apply a non-uniform magnetic field to exert a magnetic force on an atom. The magnitude of the force in the non-uniform field can be worked out from the energy: $U = -\boldsymbol{\mu} \cdot \boldsymbol{B} = -(\mu_x B_x + \mu_y B_y + \mu_z B_z)$. With $B_x = B_y = 0$ and $F_z = -\partial U/\partial z$, eqn 5.15 follows directly.

The original Stern–Gerlach experiment was performed on silver atoms, which have a ground-state electronic configuration of [Kr] $4d^{10}$ $5s^1$. Filled shells have no net orbital angular momentum, because there are as many positive m_l states occupied as negative ones. Furthermore, electrons in s-shells have $l = 0$ and therefore the orbital angular momentum of the atom is zero. This implies that the orbital magnetic dipole of the atom is also zero, and hence we expect no deflection. However, the experiment showed that the atoms were deflected either up or down, as indicated in Fig. 5.1.

In order to explain the up/down deflection of the atoms with no orbital angular momentum, we have to assume that each electron possesses an additional type of magnetic dipole moment. This magnetic dipole is attributed to the spin angular momentum. In analogy with orbital angular momentum, spin angular momentum is described by two quantum numbers s and m_s, where m_s can take the $(2s + 1)$ values in integer steps from $-s$ to $+s$. The magnitude of the spin angular momentum is given by

$$|\boldsymbol{s}| = \sqrt{s(s+1)}\hbar\,, \tag{5.16}$$

and the component along the z axis is given by

$$s_z = m_s\hbar\,. \tag{5.17}$$

The fact that atoms with a single s-shell valence electron (e.g. silver) are only deflected in two directions (i.e. up or down) implies that $(2s + 1) = 2$ and hence that $s = 1/2$. Hence the spin quantum numbers of the electron can have the following values:

$$\begin{aligned} s &= 1/2\,, \\ m_s &= \pm 1/2\,. \end{aligned}$$

The Stern–Gerlach experiment is just one of many pieces of evidence that support the hypothesis for electron spin. Here is an incomplete list of other evidence for spin based on atomic physics:

- The periodic table of elements, which is the foundation of the whole subject of chemistry, cannot be explained unless we assume that the electrons possess spin.
- High resolution spectroscopy of atomic spectral lines shows that they frequently consist of closely-spaced multiplets. This fine structure is caused by spin–orbit coupling , which can only be explained by postulating that electrons possess spin. See Chapter 7.
- If we ignore spin, we expect to observe the normal Zeeman effect when an atom is placed in an external magnetic field. However, most atoms display the *anomalous* Zeeman effect, which is a consequence of spin. See Chapter 8.
- The ratio of the magnetic dipole moment to the angular momentum is called the gyromagnetic ratio. (See Section 7.1.) The gyromagnetic ratio can be measured directly by a number of methods. In 1915, Einstein and de Haas measured the gyromagnetic ratio of iron and came up with a value twice as large as expected. They rejected this result, assigning it to experimental errors. However, we now know that the magnetism in iron is caused by the spin rather the orbital angular momentum, and so the experimental value was correct. (The electron spin g-factor is 2: see Section 7.2.) This is a salutary lesson from the history that even great physicists like Einstein and de Haas can get their error analysis wrong!

5.3 Addition of angular momentum

Having discovered that electrons have different types of angular momentum, the question now arises as to how we add them together. Let us suppose that $\boldsymbol{C}$ is the resultant of two angular momentum vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ as shown in Fig. 5.2(a), so that:

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}\,. \tag{5.18}$$

We assume for the sake of simplicity that $|\boldsymbol{A}| > |\boldsymbol{B}|$. (The argument is unaffected if $|\boldsymbol{A}| < |\boldsymbol{B}|$.) We define θ as the angle between the two vectors, as shown in figure 5.2(a).

In *classical* mechanics the angle θ can take any value from 0° to 180°. Therefore, $|\boldsymbol{C}|$ can take any value from $(|\boldsymbol{A}| + |\boldsymbol{B}|)$ to $(|\boldsymbol{A}| - |\boldsymbol{B}|)$. This is *not* the case in *quantum* mechanics, because the lengths of

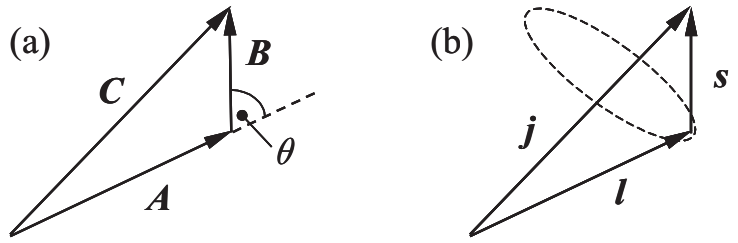

Figure 5.2: (a) Vector addition of two angular momentum vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ to form the resultant $\boldsymbol{C}$. (b) Vector model of the atom. The spin-orbit interaction couples $\boldsymbol{l}$ and $\boldsymbol{s}$ together to form the resultant $\boldsymbol{j}$. The magnitudes of the vectors are given by: $|\boldsymbol{j}| = \sqrt{j(j+1)}\hbar$, $|\boldsymbol{l}| = \sqrt{l(l+1)}\hbar$, and $|\boldsymbol{s}| = \sqrt{s(s+1)}\hbar$.

the angular momentum vectors must be quantized according to:

$$\begin{aligned} |\boldsymbol{A}| &= \sqrt{A(A+1)}\hbar \\ |\boldsymbol{B}| &= \sqrt{B(B+1)}\hbar \\ |\boldsymbol{C}| &= \sqrt{C(C+1)}\hbar\,, \end{aligned} \tag{5.19}$$

where A, B and C are the quantum numbers. This makes it apparent that θ can only take specific values in quantum mechanics. The rule for working out the allowed values of C from the known values of A and B is as follows:

$$C = A \oplus B = (A+B),\,(A+B-1),\,\cdots,\,|A-B|\,, \tag{5.20}$$

where the $\oplus$ symbol indicates that we are adding together angular momentum quantum numbers. Here are some examples of the rule given in eqn 5.20:

- $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$, $L = 3$, $S = 1$:
  $J = 3 \oplus 1$, $3 + 1 = 4$, $|3 - 1| = 2$, therefore $J = 4, 3, 2$.

- $\boldsymbol{L} = \boldsymbol{l}_1 + \boldsymbol{l}_2$, $l_1 = 2$, $l_2 = 0$:
  $L = 2 \oplus 0$, $2 + 0 = 2$, $|2 - 0| = 2$, therefore $L = 2$.

- $\boldsymbol{S} = \boldsymbol{s}_1 + \boldsymbol{s}_2$, $s_1 = 1/2$, $s_2 = 1/2$:
  $S = 1/2 \oplus 1/2$, $1/2 + 1/2 = 1$, $|1/2 - 1/2| = 0$, therefore $S = 1, 0$.

- $\boldsymbol{J} = \boldsymbol{j}_1 + \boldsymbol{j}_2$, $j_1 = 5/2$, $j_2 = 3/2$:
  $J = 5/2 \oplus 3/2$, $5/2 + 3/2 = 4$, $|5/2 - 3/2| = 1$, therefore $J = 4, 3, 2, 1$.

5.4 Spin-orbit coupling

The orbital and spin angular momenta of electrons in atoms are not totally independent of each other, but interact through the **spin-orbit interaction**. Spin-orbit coupling and its effects are considered in detail in Chapter 7, and at this stage we just need to know two basic things:

1. Spin-orbit coupling derives from the interaction between the magnetic dipole due to spin and the magnetic field that the electron experiences due to its orbital motion. We can thus write the spin-orbit interaction in the form (see eqn 7.35):

   $$\hat{H} = -\boldsymbol{\mu}_{\text{spin}} \cdot \boldsymbol{B}_{\text{orbital}} \propto \boldsymbol{l} \cdot \boldsymbol{s}\,, \tag{5.21}$$

   since $\boldsymbol{\mu}_{\text{spin}} \propto \boldsymbol{s}$ and $\boldsymbol{B}_{\text{orbital}} \propto \boldsymbol{l}$.

2. The spin-orbit interaction scales roughly as Z^2. (See eqn 7.45.) It is therefore weak in light atoms, and stronger in heavy atoms.

We introduce the spin-orbit interaction here because it is one of the mechanisms that is important in determining the angular momentum coupling schemes that apply in different atoms.

5.5 Angular momentum coupling in single electron atoms

If an atom has just a single electron, the addition of the orbital and spin angular momenta is relatively straightforward. The physical mechanism that couples the orbital and spin angular momenta together is the spin-orbit interaction, and the resultant total angular momentum vector $\boldsymbol{j}$ is defined by:

$$\boldsymbol{j} = \boldsymbol{l} + \boldsymbol{s}\,. \tag{5.22}$$

$\boldsymbol{j}$ is described by the quantum numbers j and m_j according to the usual rules for quantum mechanical angular momenta, namely:

$$|\boldsymbol{j}| = \sqrt{j(j+1)}\hbar\,, \tag{5.23}$$

and

$$j_z = m_j\hbar\,, \tag{5.24}$$

where m_j takes values of $j, (j-1), \cdots, -j$. The addition of $\boldsymbol{l}$ and $\boldsymbol{s}$ to form the resultant $\boldsymbol{j}$ is illustrated by Fig. 5.2(b).

The allowed values of j are worked out by applying eqn 5.20, with the knowledge that the spin quantum number s is always equal to $1/2$. If the electron is in a state with orbital quantum number l, we then find $j = l \oplus s = (l \pm 1/2)$, except when $l = 0$, in which case we just have $j = 1/2$. In the second case, the angular momentum of the atom arises purely from the electron spin.

5.6 Angular momentum coupling in multi-electron atoms

The Hamiltonian for an N-electron atom can be written in the form:

$$\hat{H} = \hat{H}_0 + \hat{H}_1 + \hat{H}_2\,, \tag{5.25}$$

where:

$$\hat{H}_0 = \sum_{i=1}^{N}\left(-\frac{\hbar^2}{2m}\boldsymbol{\nabla}_i^2 + V_{\text{central}}(r_i)\right)\,, \tag{5.26}$$

$$\hat{H}_1 = -\sum_{i=1}^{N}\frac{Ze^2}{4\pi\epsilon_0 r_i} + \sum_{i>j}^{N}\frac{e^2}{4\pi\epsilon_0|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \sum_{i=1}^{N}V_{\text{central}}(r_i)\,, \tag{5.27}$$

$$\hat{H}_2 = \sum_{i=1}^{N}\xi(r_i)\boldsymbol{l}_i\cdot\boldsymbol{s}_i\,. \tag{5.28}$$

As discussed in Section 4.1, $\hat{H}_0$ is the central-field Hamiltonian and $\hat{H}_1$ is the residual electrostatic potential. $\hat{H}_2$ is the spin-orbit interaction summed over the electrons of the atom.

In Chapter 4 we neglected both $\hat{H}_1$ and $\hat{H}_2$, and just concentrated on $\hat{H}_0$. This led to the conclusion that each electron occupies a state in a shell defined by the quantum numbers n and l. The reason why we neglected $\hat{H}_1$ is that the off-radial forces due to the electron-electron repulsion are smaller than the radial ones, while $\hat{H}_2$ was neglected because the spin-orbit effects are much smaller than the main terms in the Hamiltonian. It is now time to study what happens when these two terms are included. In doing so, there are two obvious limits to consider:[7]

- **LS coupling**: $\hat{H}_1 \gg \hat{H}_2$.
- **jj coupling**: $\hat{H}_2 \gg \hat{H}_1$.

Since the spin-orbit interaction scales approximately as Z^2, LS-coupling mainly occurs in atoms with small to medium Z, while jj-coupling occurs in some atoms with large Z. In the sections below, we focus on the LS-coupling limit. The less common case of jj-coupling is considered briefly in Section 5.10.

[7] In some atoms with medium-large Z (e.g. germanium $Z = 32$) we are in the awkward situation where neither limit applies. We then have **intermediate coupling**, and the behaviour is quite complicated to describe.

5.7 LS coupling

In the **LS-coupling** limit (alternatively called **Russell–Saunders coupling**), the residual electrostatic interaction is much stronger than the spin-orbit interaction. We therefore deal with the residual electrostatic interaction first and then apply the spin-orbit interaction as a perturbation. The LS coupling regime applies to most atoms of small and medium atomic number.

Let us first discuss some issues of notation. We shall need to distinguish between the quantum numbers that refer to the individual electrons within an atom and the state of the atom as a whole. The convention is:

- Lower case quantum numbers (j, l, s) refer to *individual electrons* within atoms.
- Upper case quantum numbers (J, L and S) refer to the angular momentum states of the *whole atom*.

For single electron atoms like hydrogen, there is no difference. However, in multi-electron atoms there is a real difference because we must distinguish between the angular momentum states of the individual electrons and the resultants which give the angular momentum states of the whole atom.

We can use this notation to determine the angular momentum states that the LS-coupling scheme produces. The residual electrostatic interaction has the effect of coupling the orbital and spin angular momenta of the individual electrons together, so that we find their resultants according to:

$$\boldsymbol{L} = \sum_i \boldsymbol{l}_i \,, \tag{5.29}$$

$$\boldsymbol{S} = \sum_i \boldsymbol{s}_i \,. \tag{5.30}$$

Filled shells of electrons have no net angular momentum, and so the summation only needs to be carried out over the valence electrons. In a many-electron atom, the rule given in eqn 5.20 usually allows several possible values of the quantum numbers L and S for a particular electronic configuration. Their energies will differ due to the residual electrostatic interaction. The atomic states defined by the values of L and S are called **terms**.

For each atomic term, we can find the total angular momentum of the whole atom from:

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S} \,. \tag{5.31}$$

The values of J, the quantum number corresponding to $\boldsymbol{J}$, are found from L and S according to eqn 5.20. The states of different J for each LS-term have different energies due to the spin-orbit interaction. In analogy with eqn 5.21, the spin-orbit interaction of the whole atom is written:

$$\Delta E_{\rm so} \propto -\boldsymbol{\mu}_{\rm spin}^{\rm atom} \cdot \boldsymbol{B}_{\rm orbital}^{\rm atom} \propto \boldsymbol{L} \cdot \boldsymbol{S} \,, \tag{5.32}$$

where the 'atom' superscript indicates that we take the resultant values for the whole atom. The details of the spin-orbit interaction in the LS coupling limit are considered in Section 7.6. At this stage, all we need to know is that the spin-orbit interaction splits the LS terms into **levels** labelled by J.

It is convenient to introduce a shorthand notation to label the energy levels that occur in the LS coupling regime. Each level is labelled by the quantum numbers J, L and S and is represented in the form:

$$^{2S+1}\mathrm{L}_J \,.$$

The factors $(2S+1)$ and J appear as numbers, whereas L is a letter that follows the rule:[8]

- S implies $L = 0$,
- P implies $L = 1$,
- D implies $L = 2$,
- F implies $L = 3$.

[8] The letters increment alphabetically for values of L above 3, with the exception that the letter 'J' is omitted in order to avoid confusion with the angular momentum quantum number J. Hence $L = 6$ is designated by I, but $L = 7$ is designated by K.

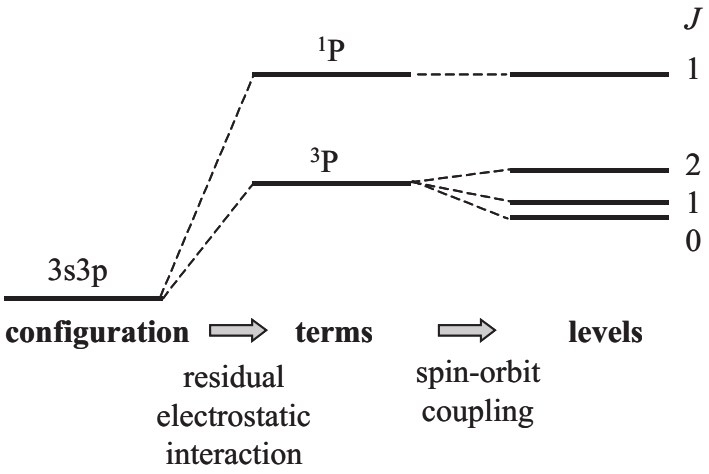

Figure 5.3: Splitting of the energy levels for the (3s,3p) configuration of magnesium in the LS coupling regime.

Thus, for example, a $^2P_{1/2}$ term is the energy level with quantum numbers $S = 1/2$, $L = 1$, and $J = 1/2$, while a $^3D_3$ has $S = 1$, $L = 2$ and $J = 3$. The factor of $(2S+1)$ in the top left is called the **multiplicity**. It indicates the degeneracy of the level due to the spin: i.e. the number of M_S states available. If $S = 0$, the multiplicity is 1, and the terms are called **singlets**. If $S = 1/2$, the multiplicity is 2 and we have **doublet** terms. If $S = 1$ we have **triplet** terms, etc.

As an example, consider the (3s,3p) electronic configuration of magnesium, where one of the valence electrons is in an s-shell with $l = 0$ and the other is in a p-shell with $l = 1$. We first work out the LS terms:

- $L = l_1 \oplus l_2 = 0 \oplus 1 = 1$.
- $S = s_1 \oplus s_2 = 1/2 \oplus 1/2 = 1$ or 0.

We thus have two terms: a ^3P triplet and a ^1P singlet. The allowed levels are then worked out as follows:

- For the ^3P triplet, we have $J = L \oplus S = 1 \oplus 1 = 2$, 1, or 0. We thus have three levels: $^3P_2$, $^3P_1$, and $^3P_0$.
- For the ^1P singlet, we have $J = L \oplus S = 1 \oplus 0 = 1$. We thus have a single $^1P_1$ level.

These levels are illustrated in Fig. 5.3. The ordering of the energy states should not concern us at this stage. The main point to realize is the general way the states split as the new interactions are turned on, and the terminology used to designate the states.

5.8 Electric dipole selection rules in the LS coupling limit

When considering electric-dipole transitions between the states of many-electron atoms that have LS-coupling, a single electron makes a jump from one atomic shell to a new one. The rules that apply to this electron are the same as the ones discussed in Section 3.4. However, we also have to think about the angular momentum state of the whole atom as specified by the quantum numbers (L, S, J). The rules that emerge are as follows:

1. The parity of the wave function must change.
2. $\Delta l = \pm 1$ for the electron that jumps between shells.
3. $\Delta L = 0, \pm 1$, but $L = 0 \to 0$ is forbidden.[9]
4. $\Delta J = 0, \pm 1$, but $J = 0 \to 0$ is forbidden.

[9] $\Delta L = 0$ transitions are obviously forbidden in one-electron atoms, because $L = l$ and l must change. However, in atoms with more than one valence electron, it is possible to get transitions between different configurations that satisfy rule 2, but have the same value of L. An example is the allowed $3p3p\,^3P_1 \to 3p4s\,^3P_2$ transition in silicon at 250.6 nm.

5. $\Delta S = 0$.

Rule 1 follows from the odd parity of the dipole operator. Rule 2 applies the $\Delta l = \pm 1$ single-electron rule to the individual electron that makes the jump in the transition, while Rule 3 applies Rule 2 to the resultant orbital angular momentum of the whole atom according to the rules for addition of angular momenta. Rule 4 follows from the fact that the total angular momentum must be conserved in the transition, allowing us to write:

$$\boldsymbol{J}^{\text{initial}} = \boldsymbol{J}^{\text{final}} + \boldsymbol{J}^{\text{photon}} . \tag{5.33}$$

The photon carries one unit of angular momentum, and so we conclude from eqn 5.20 that $\Delta J = -1,\ 0,$ or $+1$. However, the $\Delta J = 0$ rule cannot be applied to $J = 0 \to 0$ transitions because it is not possible to satisfy eqn 5.33 in these circumstances. Finally, rule 5 is a consequence of the fact that the photon does not interact with the spin.[10]

5.9 Hund's rules

We have seen above that there are many terms in the energy spectrum of a multi-electron atom. Of these, one will have the lowest energy, and will form the **ground state**. All the others are excited states. Each atom has a *unique* ground state, which is determined by minimizing the energy of its valence electrons with the residual electrostatic and spin-orbit interactions included. In principle, this is a very complicated calculation. Fortunately, however, **Hund's rules** allow us to determine which level is the ground state for atoms that have LS-coupling without lengthy calculation. The rules are:

1. The term with the largest multiplicity (i.e. largest S) has the lowest energy.
2. For a given multiplicity, the term with the largest L has the lowest energy.
3. The level with $J = |L - S|$ has the lowest energy if the shell is less than half full. If the shell is more than half full, the level with $J = L + S$ has the lowest energy.

The first of these rules basically tells us that the electrons try to align themselves with their spins parallel in order to minimize the exchange interaction. (See Chapter 6.) The other two follow from the minimizing the spin-orbit interaction.

Let us have a look at carbon as an example. Carbon has an atomic number $Z = 6$ with two valence electrons in the outermost 2p shell. Each valence electron therefore has $l = 1$ and $s = 1/2$. Consider first the (2p,np) excited state configuration with one electron in the 2p shell and the other in the np shell, where $n \geq 3$. We have from eqn 5.20 that $L = 1 \oplus 1 = 0$, 1 or 2, and $S = 1/2 \oplus 1/2 = 0$ or 1. We thus have three singlet terms (^{1}S, ^{1}P, ^{1}D), and three triplet terms (^{3}S, ^{3}P, ^{3}D). This gives rise to three singlet levels:

$$^1\mathrm{S}_0,\ \ ^1\mathrm{P}_1,\ \ ^1\mathrm{D}_2 ,$$

and seven triplet levels:

$$^3\mathrm{S}_1,\ \ ^3\mathrm{P}_0,\ \ ^3\mathrm{P}_1,\ \ ^3\mathrm{P}_2,\ \ ^3\mathrm{D}_1,\ \ ^3\mathrm{D}_2,\ \ ^3\mathrm{D}_3 .$$

We thus have a confusing array of *ten* levels in the energy spectrum for the (2p,np) configuration.

The situation in the ground state configuration (2p,2p) is simplified by the fact that the electrons are **equivalent**, i.e. in the *same* shell. The Pauli exclusion principle forbids the possibility that two or more electrons should have the same set of quantum numbers, and in the case of an atom with two valence electrons, it can be shown that this implies that $L + S$ must be equal to an even number. There is no easy explanation for this rule, but the simplest example of its application, namely to two electrons in the same s-shell, is considered in Section 6.3. For these two s-electrons, we have $L = 0 \oplus 0 = 0$ and $S = 1/2 \oplus 1/2 = 0$ or 1, giving rise to two terms: ^{1}S and ^{3}S. Both terms are allowed when the electrons are in different s-shells, but the $L + S =$ even rule tells us that only the singlet ^{1}S term is allowed if the electrons are in the same s-shell. The proof that the triplet term does not exist for the (1s,1s) ground-state configuration of helium is given in Section 6.3.

On applying the rule that $L + S$ must be even to the equivalent 2p electrons in the carbon ground state, we find that only the ^{1}S, ^{1}D, and ^{3}P terms are allowed, which means that only five of the ten levels listed above are possible:[11]

$$^1\mathrm{S}_0,\ \ ^1\mathrm{D}_2,\ \ ^3\mathrm{P}_0,\ \ ^3\mathrm{P}_1,\ \ ^3\mathrm{P}_2 .$$

[10] $\Delta S \neq 0$ transitions can be weakly allowed when the spin-orbit coupling is strong, because the spin is then mixed with the orbital motion.

[11] The full derivation of the allowed states for the (np,np) configuration of a group IV atom is considered, for example, in Woodgate, *Elementary Atomic Structure*, 2nd Edition, Oxford University Press, 1980, Section 7.2.

	m_l		
m_s	–1	0	+1
+1/2		↑	↑
−1/2			

Table 5.1: Distribution of the two valence electrons of the carbon ground state within the m_s and m_l states of the 2p shell.

We can now apply Hund's rules to find out which of these is the ground state. The first rule states that the triplet levels have the lower energy. Since these all have $L = 1$ we do not need to consider the second rule. The shell is less than half full, and so we have $J = |L - S| = 0$. The ground state is thus the $^3\mathrm{P}_0$ level. All the other levels are excited states.

It is important to notice that, if we had forgotten the rule that $L + S$ must be even, we would have incorrectly concluded from Hund's rules that the ground state is a $^3\mathrm{D}_1$ term, which does not exist for the (2p,2p) configuration. It is therefore safer to use a different version of Hund's rules, based on the allowed combinations of (m_s, m_l) sub-levels:

1. Maximize the spin and set $S = \sum m_s$.
2. Maximize the orbital angular momentum, subject to rule 1, and set $L = \sum m_l$.
3. $J = |L - S|$ if the shell is less than half full, otherwise $J = |L + S|$.

These rules should work in all cases, since they incorporate the Pauli exclusion principle properly.

As an example of how to use the second version of Hund's rules, we apply them again to the two 2p electrons of carbon. The two electrons can go into the six possible (m_s, m_l) sub-levels of the 2p shell.

1. To get the largest value of the spin, we must have both electron spins aligned with $m_s = +1/2$. This gives $S = 1/2 + 1/2 = 1$.
2. Having put both electrons into spin up states, we cannot now put both electrons into $m_l = +1$ states because of Pauli's exclusion principle. The best we can do is to put one into an $m_l = 1$ state and the other into an $m_l = 0$ state, as illustrated in Table 5.1. This gives $L = 1 + 0 = 1$.
3. The shell is less than half full, and so we have $J = |L - S| = 0$.

We thus deduce that the ground state is the $^3\mathrm{P}_0$ level, as before.

The ground state levels for the first 11 elements, as worked out from Hund's rules, are listed in Table 5.2. Experimental results confirm these predictions. Note that full shells always give $^1\mathrm{S}_0$ level with no net angular momentum: $S = L = J = 0$.

It is important to be aware that Hund's rules *cannot* be used to find the energy ordering of excited states with reliability. For example, consider the (2p,3p) excited state configuration of carbon, which has the ten possible levels listed previously. Hund's rules predict that the $^3\mathrm{D}_1$ level has the lowest energy, but the lowest state is actually the $^1\mathrm{P}_1$ level.

5.10 jj coupling

The spin-orbit interaction gets larger as Z increases. (See, for example, eqn 7.45.) This means that in some atoms with large Z (eg tin with $Z = 50$) we can have a situation in which the spin-orbit interaction is much stronger than the residual electrostatic interaction. In this regime, **jj coupling** coupling occurs. The spin-orbit interaction couples the orbital and spin angular momenta of the individual electrons together first, and we then find the resultant $\boldsymbol{J}$ for the whole atom by adding together the individual $\boldsymbol{j}$s:

$$\begin{aligned} \boldsymbol{j}_i &= \boldsymbol{l}_i + \boldsymbol{s}_i \\ \boldsymbol{J} &= \sum_{i=1}^{N} \boldsymbol{j}_i \end{aligned} \tag{5.34}$$

These J states are then split by the weaker residual electrostatic potential, which acts as a perturbation.

Z	Element	Configuration	Ground state
1	H	$1s^1$	$^2S_{1/2}$
2	He	$1s^2$	$^1S_0$
3	Li	$1s^2\ 2s^1$	$^2S_{1/2}$
4	Be	$1s^2\ 2s^2$	$^1S_0$
5	B	$1s^2\ 2s^2\ 2p^1$	$^2P_{1/2}$
6	C	$1s^2\ 2s^2\ 2p^2$	$^3P_0$
7	N	$1s^2\ 2s^2\ 2p^3$	$^4S_{3/2}$
8	O	$1s^2\ 2s^2\ 2p^4$	$^3P_2$
9	F	$1s^2\ 2s^2\ 2p^5$	$^2P_{3/2}$
10	Ne	$1s^2\ 2s^2\ 2p^6$	$^1S_0$
11	Na	$1s^2\ 2s^2\ 2p^6\ 3s^1$	$^2S_{1/2}$

Table 5.2: Electronic configurations and ground state terms of the first 11 elements in the periodic table.

Reading

Bransden and Joachain, *Physics of Atoms and Molecules*, §1.8, 2.5, 8.5, 9.2
Demtröder, W., *Atoms, Molecules and Photons*, §5.5–6, 6.2–5.
Haken and Wolf, *The physics of atoms and quanta*, chapters 12, 17, 19.
Hertel and Schulz, *Atoms, Molecules and Optical Physics, 1*, §1.9 – 1.10, 2.5, 6.2, 10.4
Eisberg and Resnick, *Quantum Physics*, chapters 8, 10.
Foot, *Atomic physics*, §2.3.1, chapter 5.
Beisser, *Concepts of Modern Physics*, §7.7 – 8.

Chapter 6

Helium and exchange symmetry

In this chapter we will look at atoms with two valence electrons. This includes helium, and the group II elements: beryllium, magnesium, calcium, etc. As we will see, this leads to the idea of the exchange energy. We shall use helium as the main example, as it is a true two electron system and illustrates the physical points most clearly.

6.1 Exchange symmetry

Consider a multi-electron atom with N electrons, as illustrated in figure 6.1(a). The wave function of the atom will be a function of the co-ordinates of the individual electrons:

$$\Psi \equiv \Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_K, \boldsymbol{r}_L, \cdots \boldsymbol{r}_N)$$

However, the electrons are **indistinguishable** particles. It is not physically possible to stick labels on the individual electrons and then keep tabs on them as the move around their orbits. This means that the many-electron wave function must have **exchange symmetry**:

$$|\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_K, \boldsymbol{r}_L, \cdots \boldsymbol{r}_N)|^2 = |\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_L, \boldsymbol{r}_K, \cdots \boldsymbol{r}_N)|^2 \,. \tag{6.1}$$

This says that nothing happens if we switch the labels of any pair of electrons. Equation 6.1 will be satisfied if

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_K, \boldsymbol{r}_L, \cdots \boldsymbol{r}_N) = \pm\, \Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_L, \boldsymbol{r}_K, \cdots \boldsymbol{r}_N)\,. \tag{6.2}$$

The + sign in equation 6.2 applies if the particles are **bosons**. These are said to be **symmetric** with respect to particle exchange. The − sign applies to **fermions**, which are **anti-symmetric** with respect to particle exchange.

Electrons have spin 1/2 and are therefore fermions. Hence the wave function of a multi-electron atom must be *anti-symmetric* with respect to particle exchange. This is a very fundamental property, and is the physical basis of the Pauli exclusion principle, as we shall see below.

The discussion of exchange symmetry gets quite complicated when there are lots of electrons, and so we shall just concentrate on helium here.

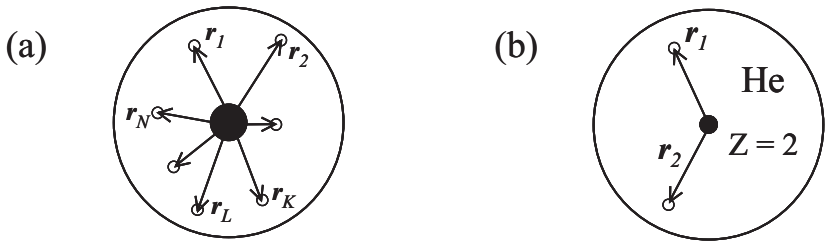

Figure 6.1: (a) A multi-electron atom with N electrons. (b) The helium atom.

ψ_{spatial}	ψ_{spin}
symmetric	anti-symmetric ($S = 0$)
anti-symmetric	symmetric ($S = 1$)

Table 6.1: Allowed combinations of the exchange symmetries of the spatial and spin wave functions of fermionic particles.

6.2 Helium wave functions

Figure 6.1(b) shows a schematic diagram of a helium atom. It consists of one nucleus with $Z = 2$ and two electrons. The position co-ordinates of the electrons are written $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ respectively.

The quantum state in the helium atom will be specified both by the spatial co-ordinates and by the spin of the two electrons. The two-electron wave function is therefore written as a product of a spatial wave function and a spin wave function:

$$\Psi = \psi_{\text{spatial}}(\boldsymbol{r}_1, \boldsymbol{r}_2)\, \psi_{\text{spin}} \,. \tag{6.3}$$

As we have seen above, the fact that electrons are indistinguishable fermions requires that the two-electron wave function Ψ must be anti-symmetric with respect to exchange of electrons 1 and 2. Table 6.1 lists the two possible combinations of wave function symmetries that can produce an antisymmetric total wave function.

We first consider the spin wave function. We have two spin 1/2 electrons, and so the total spin quantum number S is given by $S = 1/2 \oplus 1/2 = 1$ or 0. $S = 0$ states are called **singlets** because they only have one possible M_S value, namely 0, while $S = 1$ states are called **triplets** because they have three possible M_S values, namely $+1$, 0, and -1.

There are four possible ways of combining the spins of the two electrons so that the total wave function has exchange symmetry. These are listed in Table 6.2. The component of $\boldsymbol{S}$ along the z-axis is obtained by adding together the s_z values of the individual electrons. This gives the S_z value of the whole helium atom, and hence the spin quantum number M_S.

Inspection of Table 6.2 shows us that we have three symmetric spin states with M_S values of $+1$, 0 and -1 and one antisymmetric spin wave function with $M_S = 0$. The $M_S = +1$ and $M_S = -1$ wave functions are derived unambiguously from the triplet state. On the other hand, the two $M_S = 0$ wave functions could come from either the singlet or triplet states. However, the spin states must have well-defined exchange symmetries, and the $M_S = \pm 1$ wave functions are clearly symmetric. This implies that the symmetric $M_S = 0$ wave function comes from the triplet state, and hence that the anti-symmetric wave function corresponds to the singlet state. We thus conclude that triplet states have symmetric spin wave functions, while singlets have anti-symmetric spin wave functions.

Now let us consider the spatial wave functions. The state of the atom will be specified by the configuration of the two electrons. In the ground state both electrons are in the 1s shell, and so we have a configuration of $1s^2$. In the excited states, one or both of the electrons will be in a higher shell. The configuration is thus given by the n, l values of the two electrons, and we write the configuration as $(n_1 l_1,\ n_2 l_2)$. This means that the spatial part of the helium wave function must contain terms of the type $u_A(\boldsymbol{r}_1)\, u_B(\boldsymbol{r}_2)$, where $u_{nl}(\boldsymbol{r})$ is the wave function for an electron with quantum numbers n and l, and the subscripts A and B stand for the quantum numbers n, l of the two electrons.

The discussion above does not take account of the fact that the electrons are indistinguishable: we cannot distinguish between the state with electron 1 in state A and electron 2 in state B, and *vice versa*. $u_B(\boldsymbol{r}_1)\, u_A(\boldsymbol{r}_2)$ is therefore an equally valid wave function for the particular electronic configuration. The wave function for the configuration A, B must therefore take the form:

$$\psi_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{1}{\sqrt{2}}\Big(u_A(\boldsymbol{r}_1)\, u_B(\boldsymbol{r}_2) \pm u_B(\boldsymbol{r}_1)\, u_A(\boldsymbol{r}_2)\Big) \,. \tag{6.4}$$

The $1/\sqrt{2}$ factor ensures that $\psi_{AB}(\boldsymbol{r}_1, \boldsymbol{r}_2)$ is correctly normalized. It is easy to verify that the wave function with the $+$ sign is symmetric with respect to particle exchange, while the wave function with the $-$ sign is antisymmetric.

Spin wave function	symmetry	M_S
$\uparrow_1 \uparrow_2$	+	+1
$\frac{1}{\sqrt{2}}(\uparrow_1 \downarrow_2 + \downarrow_1 \uparrow_2)$	+	0
$\frac{1}{\sqrt{2}}(\uparrow_1 \downarrow_2 - \downarrow_1 \uparrow_2)$	−	0
$\downarrow_1 \downarrow_2$	+	−1

Table 6.2: Spin wave functions for a two-electron system. The arrows indicate whether the spin of the individual electrons is up or down (ie $+\frac{1}{2}$ or $-\frac{1}{2}$). The + sign in the symmetry column applies if the wave function is symmetric with respect to particle exchange, while the − sign indicates that the wave function is anti-symmetric. The S_z value is indicated by the quantum number for M_S, which is obtained by adding the m_s values of the two electrons together.

S	M_S	ψ_{spin}	ψ_{spatial}
0	0	$\frac{1}{\sqrt{2}}(\uparrow_1 \downarrow_2 - \downarrow_1 \uparrow_2)$	$\frac{1}{\sqrt{2}}\Big(u_A(\boldsymbol{r}_1)\,u_B(\boldsymbol{r}_2) + u_B(\boldsymbol{r}_1)\,u_A(\boldsymbol{r}_2)\Big)$
1	+1	$\uparrow_1 \uparrow_2$	$\frac{1}{\sqrt{2}}\Big(u_A(\boldsymbol{r}_1)\,u_B(\boldsymbol{r}_2) - u_B(\boldsymbol{r}_1)\,u_A(\boldsymbol{r}_2)\Big)$
	0	$\frac{1}{\sqrt{2}}(\uparrow_1 \downarrow_2 + \downarrow_1 \uparrow_2)$	
	−1	$\downarrow_1 \downarrow_2$	

Table 6.3: Spin and spatial wave functions for a two-electron atom with electronic configuration designated by the labels A and B.

We have seen above that spin singlet and triplet states are, respectively, antisymmetric and symmetric under exchange symmetry. The fact that the overall symmetry must be negative, as summarized in Table 6.1, then implies that spin singlets and triplets must be paired off with symmetric and antisymmetric spatial wave functions respectively. This leads to the detailed pairing of spin and spatial wave functions shown in Table 6.3. The key point is that the singlet and triplet states have different spatial wave functions, which has a strong effect on the energy of the atom, as we shall see below. This is a surprising result when you consider that the spin and spatial co-ordinates are basically independent of each other.

6.3 The Pauli exclusion principle

Let us suppose that we try to put the two electrons in the same atomic shell. The ground state of helium is an example of such a configuration, with both electrons in the 1s shell. The spatial wave functions will be given by eqn 6.4 with $A = B$. The antisymmetric combination with the − sign in the middle is zero in this case. From Table 6.3 we see that this implies that there are no triplet $S = 1$ states if both electrons are in the same shell.

The fact that the triplet state does not exist for the helium ground state is a demonstration of the rule that $L + S$ must be even for a two-electron atom with both electrons in the same shell. In the case of the $1s^2$ configuration, we have $L = 0$, and therefore $S = 1$ is not allowed. This rule was introduced without any justification in Section 5.9. The general justification of the rule is beyond the scope of this course, but the example of the helium ground state at least demonstrates that the rule is true for the simplest case.

The absence of the triplet state for $1s^2$ configuration is equivalent to the **Pauli exclusion principle**.

We are trying to put two electrons in the same state as defined by the n, l, m_l quantum numbers. This is only possible if the two electrons have different m_s values. In other words, their spins must be aligned anti-parallel. The $S = 1$ state contains terms with both spins pointing in the same direction, and is therefore not allowed. The analysis of the symmetry of the wave function discussed here thus shows us that the Pauli exclusion principle is a consequence of the fact that electrons are indistinguishable fermions.

6.3.1 Slater determinants

We note in passing that the anti-symmetric wave function given in eqn. 6.4 can be written as a determinant:

$$\psi_{\rm spatial} = \frac{1}{\sqrt{2}} \begin{vmatrix} u_A(\boldsymbol{r}_1) & u_A(\boldsymbol{r}_2) \\ u_B(\boldsymbol{r}_1) & u_B(\boldsymbol{r}_2) \end{vmatrix} . \tag{6.5}$$

This can be generalized to give the correct anti-symmetric wave function when we have more than two electrons:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} u_\alpha(1) & u_\alpha(2) & \cdots & u_\alpha(N) \\ u_\beta(1) & u_\beta(2) & \cdots & u_\beta(N) \\ \vdots & \vdots & \ddots & \vdots \\ u_\nu(1) & u_\nu(2) & \cdots & u_\nu(N) \end{vmatrix} , \tag{6.6}$$

where $\{\alpha, \beta, \cdots, \nu\}$ each represent a set of quantum numbers $\{n, l, m_l, m_s\}$ for the individual electrons, and $\{1, 2, \cdots, N\}$ are the electron labels. Determinants of this type are called **Slater determinants**. Note that the determinant is zero if any two rows are equal, which tells us that each electron in the atom must have a unique set of quantum numbers, as required by the Pauli exclusion principle.

We shall not make further use of Slater determinants in this course. They are mentioned here for completeness.

6.4 The exchange energy

The Hamiltonian for the helium atom before we consider fine-structure effects is given by:

$$\hat{H} = \left(-\frac{\hbar^2}{2m}\nabla_1^2 - \frac{2e^2}{4\pi\epsilon_0 r_1} \right) + \left(-\frac{\hbar^2}{2m}\nabla_2^2 - \frac{2e^2}{4\pi\epsilon_0 r_2} \right) + \frac{e^2}{4\pi\epsilon_0 r_{12}} , \tag{6.7}$$

where $r_{12} = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$. The first two terms enclosed in brackets account for the kinetic energy of the two electrons and their attraction towards the nucleus, which has a charge of $+2e$. The final term is the Coulomb repulsion between the two electrons. It is this Coulomb repulsion which makes the equations difficult to deal with.

In § 4.1 and following we described how to deal with a many-electron Hamiltonian by splitting it into a central field and a residual electrostatic interaction. In the case of helium, we just have one Coulomb repulsion term and it is easier to go back to first principles. We can then use the correctly symmeterized wave functions to calculate the energies for specific electronic configurations of the helium atom.

The energy of the electronic configuration $(n_1 l_1,\ n_2 l_2)$ is found by computing the expectation value of the Hamiltonian:

$$\langle E \rangle = \iint \psi^*_{\rm spatial}\, \hat{H}\, \psi_{\rm spatial}\, {\rm d}^3\boldsymbol{r}_1 {\rm d}^3\boldsymbol{r}_2 . \tag{6.8}$$

The spin wave functions do not appear here because the Hamiltonian does not affect the spin directly, and so the spin wave functions just integrate out to unity.

We start by re-writing the Hamiltonian given in eqn 6.7 in the following form:

$$\hat{H} = \hat{H}_1 + \hat{H}_2 + \hat{H}_{12} , \tag{6.9}$$

where

$$\hat{H}_i = -\frac{\hbar^2}{2m}\nabla_i^2 - \frac{2e^2}{4\pi\varepsilon_0 r_i} , \tag{6.10}$$

$$\hat{H}_{12} = \frac{e^2}{4\pi\varepsilon_0 \left|\boldsymbol{r}_1 - \boldsymbol{r}_2\right|} . \tag{6.11}$$

The energy can be split into three parts:

$$E = E_1 + E_2 + E_{12}\,, \tag{6.12}$$

where:

$$E_i = \iint \psi^*_{\rm spatial} \hat{H}_i \psi_{\rm spatial} {\rm d}^3\boldsymbol{r}_1 {\rm d}^3\boldsymbol{r}_2\,, \tag{6.13}$$

and

$$E_{12} = \iint \psi^*_{\rm spatial} \hat{H}_{12} \psi_{\rm spatial} {\rm d}^3\boldsymbol{r}_1 {\rm d}^3\boldsymbol{r}_2\,. \tag{6.14}$$

The first two terms in eqn 6.12 represent the energies of the two electrons in the absence of the electron-electron repulsion. These are just equal to the hydrogenic energies of each electron:

$$E_1 + E_2 = -\frac{4R_H}{n_1^2} - \frac{4R_H}{n_2^2}\,, \tag{6.15}$$

where the factor of $4 \equiv Z^2$ accounts for the nuclear charge. (See Appendix C for the evaluation of the integrals.) The third term is the electron-electron Coulomb repulsion energy:

$$E_{12} = \iint \psi^*_{\rm spatial} \frac{e^2}{4\pi\epsilon_0 r_{12}} \psi_{\rm spatial}\, {\rm d}^3\boldsymbol{r}_1 {\rm d}^3\boldsymbol{r}_2\,. \tag{6.16}$$

As shown in Appendix C, the end result for the correctly symmeterized wave functions given in eqn 6.4 is:

$$E_{12} = D_{AB} \pm J_{AB}, \tag{6.17}$$

where the + sign is for singlets and the − sign is for triplets. D_{AB} is the **direct** Coulomb energy given by:

$$D_{AB} = \frac{e^2}{4\pi\epsilon_0} \iint u^*_A(\boldsymbol{r}_1)\, u^*_B(\boldsymbol{r}_2)\, \frac{1}{r_{12}}\, u_A(\boldsymbol{r}_1)\, u_B(\boldsymbol{r}_2)\, {\rm d}^3\boldsymbol{r}_1\, {\rm d}^3\boldsymbol{r}_2\,, \tag{6.18}$$

and J_{AB} is the **exchange** Coulomb energy given by

$$J_{AB} = \frac{e^2}{4\pi\epsilon_0} \iint u^*_A(\boldsymbol{r}_1)\, u^*_B(\boldsymbol{r}_2)\, \frac{1}{r_{12}}\, u_B(\boldsymbol{r}_1)\, u_A(\boldsymbol{r}_2)\, {\rm d}^3\boldsymbol{r}_1\, {\rm d}^3\boldsymbol{r}_2\,. \tag{6.19}$$

Note that in the exchange integral, we are integrating the expectation value of $1/r_{12}$ with each electron in a different shell. This is why it is called the "exchange" energy. The total energy of the configuration $(n_1 l_1,\ n_2 l_2)$ is thus given by:

$$E(n_1 l_1,\ n_2 l_2) = -\frac{4R_H}{n_1^2} - \frac{4R_H}{n_2^2} + D_{AB} \pm J_{AB}\,, \tag{6.20}$$

where the + sign applies to singlet ($S = 0$) states and the − sign to triplets ($S = 1$). We thus see that the energies of the singlet and triplet states differ by $2J_{AB}$. This splitting of the spin states is a direct consequence of the exchange symmetry.

Note that:

- The exchange splitting is *not* a small energy. It is part of the gross structure of the atom. This contrasts with the other spin-dependent effect that we have considered, namely the spin-orbit interaction, which is a small relativistic correction and only contributes to the "fine" structure of the atom. The value of $2J_{AB}$ for the first excited state of helium, namely the 1s2s configuration, is 0.80 eV.

- We can give a simple physical reason why the symmetry of the spatial wave function (and hence the spin) affects the energy so much. If we put $\boldsymbol{r}_1 = \boldsymbol{r}_2$ into eqn 6.4, we see that we get $\psi_{\rm spatial} = 0$ for the anti-symmetric state. This means that the two electrons have a low probability of coming close together in the triplet state, and hence reduces their Coulomb repulsion energy. On the other hand, $\psi_{\rm spatial}(\boldsymbol{r}_1 = \boldsymbol{r}_2) \neq 0$ for singlet states with symmetric spatial wave functions. They therefore have a larger Coulomb repulsion energy.

- The exchange energy is sometimes written in the form

$$\Delta E_{\rm exchange} \propto -J\, \boldsymbol{s}_1 \cdot \boldsymbol{s}_2\,. \tag{6.21}$$

  This emphasizes the point that the change of energy is related to the relative alignment of the electron spins. If both spins are aligned, as they are in the triplet states, the energy goes down. If the spins are anti-parallel, the energy goes up.

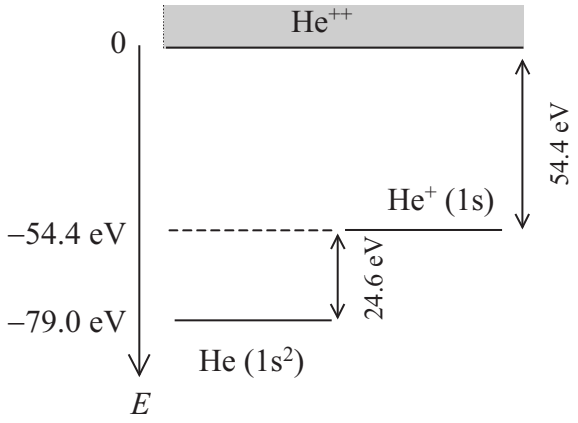

Figure 6.2: The ionization energies of helium atom.

- The notation given in eqn 6.21 is extensively used when explaining the phenomenon of **ferromagnetism** in the subject of magnetism. The energy that induces the spins to align parallel to each other is caused by the spin-dependent change of the Coulomb repulsion energy of the electrons. The magnetic energy of the electrons due to the dipole-dipole interaction is completely negligible on this scale.

6.5 The helium term diagram

The term diagram for helium can be worked out if we can evaluate the direct and exchange Coulomb energies. The total energy for each configuration is given by eqn 6.20.

The ground state

In the ground state both electrons are in the 1s shell, and so we have a configuration of $1s^2$. We have seen above that we can only have $S = 0$ for this configuration. The energy is thus given by:

$$\begin{aligned} E(1s^2) &= -\frac{4R_H}{1^2} - \frac{4R_H}{1^2} + \left(D_{1s^2} + J_{1s^2}\right) \\ &= -54.4\,\text{eV} - 54.4\,\text{eV} + 29.8\,\text{eV} \\ &= -79.0\,\text{eV}\ . \end{aligned} \tag{6.22}$$

The computation of the direct and exchange energies is non-trivial (to say the least) and keeps theoretical atomic physicists busy. The value of 29.8 eV given here can be deduced experimentally from the first ionization potential (see below).

Ionization potentials

The excited states are made by promoting one of the electrons to higher shells. When the second electron has been promoted into the energy continuum at $n_2 = \infty$, we are left with a singly ionized helium atom: He^+. This is now a hydrogenic system. We have one electron in the 1s shell orbiting around a nucleus with charge $+2e$, and the energy is just $-Z^2R_H = -54.4\,\text{eV}$. We thus deduce that the first ionization potential of helium is $-54.4 - (-79.0) = 24.6\,\text{eV}$. The second ionization potential (ie the energy required to liberate the second electron) is then equal to 54.4 eV. This point is illustrated in Fig. 6.2. Note that this is an example of the point made in the discussion of Fig. 1.5 in Section 1.3, namely that the ionization limit of the neutral He atom corresponds to the ground state of the He^+ ion.

Optical spectra

The first few excited states of helium are listed in Table 6.4. We do not need to consider "two electron jump" excited states such as the 2s2s configuration here. This is because the Bohr model tells us that

Ground state	1s 1s ($\equiv$ 1s^2)
First excited state	1s 2s
Second excited state	1s 2p
Third excited state	1s 3s
Fourth excited state	1s 3p
$\vdots$	
Ionization limit	1s ∞l

Table 6.4: Electron configurations for the states of the helium atom.

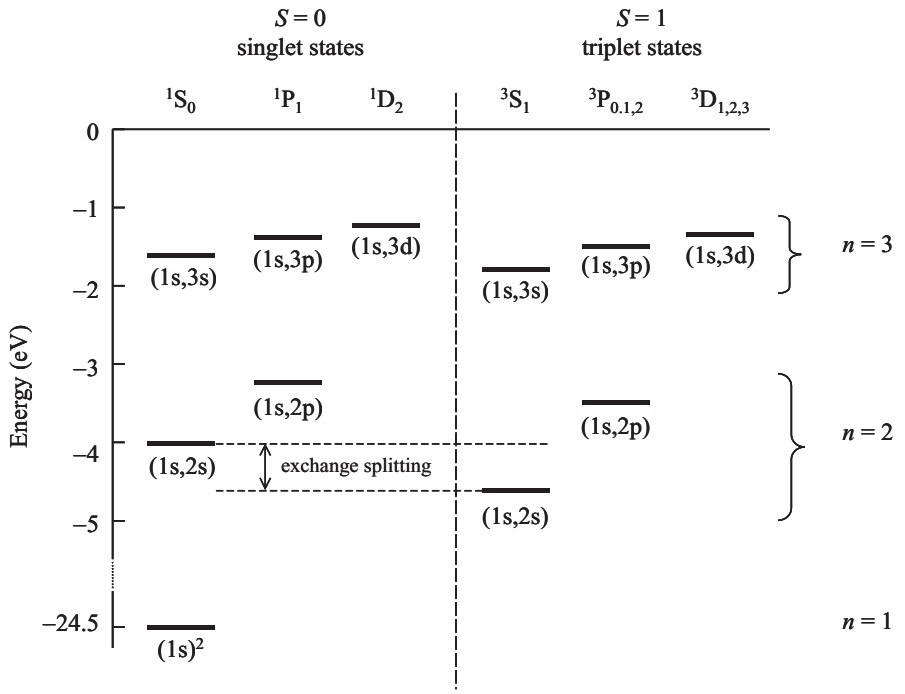

Figure 6.3: Approximate energy term diagram for helium. The diagram is split into singlet and triplet states because only $\Delta S = 0$ transitions are allowed by the selection rules. The energy difference between the singlet and triplet terms for the same configuration is caused by the exchange energy, as identified for the 1s2s configuration.

we need an energy of about $2 \times \frac{3}{4} R_H$ to promote two electrons to the $n = 2$ shell. This is larger than the first ionization energy.

For each excited state we have two spin states corresponding to S equal to 0 or 1. The triplet $S = 1$ terms are at lower energy than the singlets due to the exchange energy. (See eqn 6.17.) The $\Delta S = 0$ selection rule tells us that we cannot get optical transitions between the singlets and triplet terms. The transitions involving singlet states have a normal Zeeman effect since $S = 0$, but the triplet transitions have an anomalous Zeeman effect since $S \neq 0$.

The energy term diagram for the first few excited states are shown in Fig. 6.3. The energy of the (1s, nl) state approaches the hydrogenic energy $-R_H/n^2$ when n is large. This is because the excited electron is well outside the 1s shell, which just partly screens the nuclear potential. The outer electron just sees $Z_{\rm eff} = 1$, and we have a hydrogenic potential.

Excited states states such as the 1s 2s configuration are said to be **metastable**. They cannot relax easily to the ground state. The relaxation would involve a 2s $\rightarrow$ 1s transition, which is forbidden by the $\Delta l = \pm 1$ selection rule. Furthermore, the relaxation of the triplet 1s 2s configuration is further forbidden by the $\Delta S = 0$ selection rule. These states therefore have very long lifetimes.

6.6 Optical spectra of group II elements

The principles that we have been discussing here with respect to helium apply equally well to other two-electron atoms. In particular, they apply to the elements in group IIA of the periodic table (e.g. Be, Mg, Ca.) These atoms have two valence electrons in an s-shell outside a filled shell. The term diagram for group IIA elements would appear generically similar to Fig. 6.3, and the optical spectra would follow similar rules, with singlet and triplet transitions split by the exchange energy. The singlet and triplet transitions have normal and anomalous Zeeman effects, respectively.

Reading

Bransden and Joachain, *Physics of Atoms and Molecules*, chapter 7, §9.5
Demtröder, W., *Atoms, Molecules and Photons*, section 6.1.
Haken and Wolf, *The physics of atoms and quanta*, chapters 17 and 19.
Hertel and Schulz, *Atoms, Molecules and Optical Physics, 1*, chapter 7
Foot, *Atomic physics*, Chapter 3.
Eisberg and Resnick, *Quantum Physics*, chapter 9.
Beisser, *Concepts of Modern Physics*, chapter 7.

Chapter 7

Fine structure and nuclear effects

Up to this point, we have been mainly studying the **gross structure** of atoms. When we consider the gross structure, we include only the largest interaction terms in the Hamiltonian, namely, the electron kinetic energy, the electron-nuclear attraction, and the electron-electron repulsion.

It is now time to start considering the smaller interactions in the atom that arise from magnetic effects. In this chapter we shall consider only those effects caused by *internal* magnetic fields, leaving the discussion of the effects produced by *external* fields to the next set of notes. The internal fields within atoms cause **fine structure** in atomic spectra. We shall start by considering the fine structure of hydrogen and then move on to many-electron atoms. At the end of these notes we shall also look briefly at **hyperfine structure**, which is a similar, but smaller, effect due to the magnetic interactions between the electrons and the nucleus.

7.1 Orbital magnetic dipoles

The quantum numbers n and l were first introduced in the old quantum theory of Bohr and Sommerfeld. The **principal quantum number** n was introduced in the Bohr model as a fundamental postulate concerning the quantization of the angular momentum (see eqn 2.1), while the **orbital quantum number** l was introduced a few years later by Sommerfeld as a patch-up to account for the possibility that the atomic orbits might be elliptical rather than circular. In Section 2.2.3 we saw how these two quantum numbers naturally re-appear in the full quantum mechanical treatment of the hydrogen atom. Then, in Section 4.1, we saw how they carry across to many-electron atoms.

Two key results that drop out of the quantum mechanical treatment of atoms are:

- The magnitude L of the orbital angular momentum of an electron is given by (see eqn 2.37):

$$L = \sqrt{l(l+1)}\hbar\,, \tag{7.1}$$

  where l can take integer values up to $(n-1)$.

- The component of the angular momentum along a particular axis (usually taken as the z axis) is quantized in units of $\hbar$ and its value is given by (see eqn 2.38):

$$L_z = m_l\hbar\,, \tag{7.2}$$

  where the magnetic quantum number m_l can take integer values from $-l$ to $+l$.

These two relationships give rise to the vector model of angular momentum illustrated in Fig. 2.3.

The orbital motion of the electron causes it to have a magnetic moment. Let us first consider an electron in a circular Bohr orbit, as illustrated in Fig. 7.1(a). The electron orbit is equivalent to a current loop, and we know from electromagnetism that current loops behave like magnets. The electron in the Bohr orbit is equivalent to a little magnet with a magnetic dipole moment μ given by:

$$\mu = i \times \text{Area} = -(e/T) \times (\pi r^2)\,, \tag{7.3}$$

where T is the period of the orbit. Now $T = 2\pi r/v$, and so we obtain

$$\mu = -\frac{ev}{2\pi r}\pi r^2 = -\frac{e}{2m_\text{e}}m_\text{e}vr = -\frac{e}{2m_\text{e}}L, \tag{7.4}$$

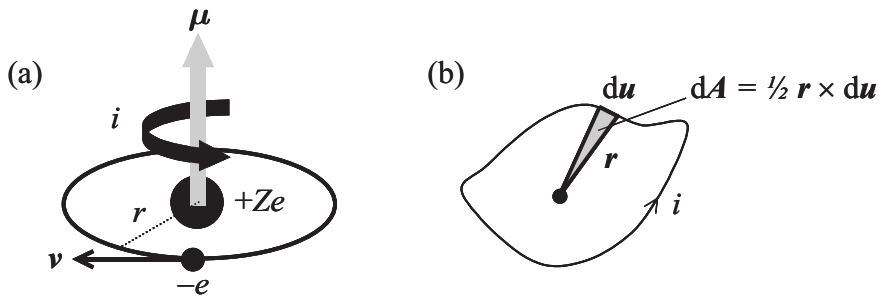

Figure 7.1: (a) The orbital motion of the electron around the nucleus in a circular Bohr orbit is equivalent to a current loop, which generates a magnetic dipole moment. (b) Magnetic dipole moment of an electron in a non-circular orbit.

where we have substituted L for the orbital angular momentum $m_e vr$.

This relationship can easily be generalized to the case of electrons in non-circular orbits. Consider an electron at position vector $\boldsymbol{r}$ in a non-circular orbit with an origin O. The magnetic dipole moment is given by:

$$\boldsymbol{\mu} = \oint i \, \mathrm{d}\boldsymbol{A} \,, \tag{7.5}$$

where i is the current in the loop and $\mathrm{d}\boldsymbol{A}$ is the incremental area swept out by the electron as it performs its orbit. The incremental area $\mathrm{d}\boldsymbol{A}$ is related to the path element $\mathrm{d}\boldsymbol{u}$ by:

$$\mathrm{d}\boldsymbol{A} = \frac{1}{2} \boldsymbol{r} \times \mathrm{d}\boldsymbol{u} \,, \tag{7.6}$$

and so eqn 7.5 becomes:

$$\boldsymbol{\mu} = \frac{1}{2} \oint i \, \boldsymbol{r} \times \mathrm{d}\boldsymbol{u} \,. \tag{7.7}$$

We can write the current as $i = \mathrm{d}q/\mathrm{d}t$, where q is the charge, which implies:

$$\begin{aligned} \boldsymbol{\mu} &= \frac{1}{2} \oint \frac{\mathrm{d}q}{\mathrm{d}t} \boldsymbol{r} \times \mathrm{d}\boldsymbol{u} \,, \\ &= \frac{1}{2} \oint \mathrm{d}q \, \boldsymbol{r} \times \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} \,, \\ &= \frac{1}{2} \oint \mathrm{d}q \, \boldsymbol{r} \times \boldsymbol{v} \,, \\ &= \frac{1}{2m_e} \oint \mathrm{d}q \, \boldsymbol{r} \times \boldsymbol{p} \,, \end{aligned} \tag{7.8}$$

where $\boldsymbol{v}$ is the velocity, and $\boldsymbol{p}$ is the momentum. The angular momentum is defined as usual by

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p} \tag{7.9}$$

and so we finally obtain:

$$\boldsymbol{\mu} = \frac{1}{2m_e} \oint \boldsymbol{L} \mathrm{d}q = \frac{1}{2m_e} \boldsymbol{L} \oint \mathrm{d}q = \frac{1}{2m_e} \boldsymbol{L}(-e) \,, \tag{7.10}$$

as in eqn 7.4. Note that the result works because the angular momentum $\boldsymbol{L}$ is a constant of the motion in the central field approximation (see Section 5.2.1), and so it can be taken out of the integral.

Equation 7.4 shows us that the orbital angular momentum is directly related to the magnetic dipole moment. The quantity $e/2m_e$ that appears is called the **gyromagnetic ratio**. It specifies the proportionality constant between the angular momentum of an electron and its magnetic moment. It is apparent from eqns 7.1 and 7.4 that the magnitude of atomic magnetic dipoles is given by:

$$|\boldsymbol{\mu}| = \frac{e}{2m_e} \hbar \sqrt{l(l+1)} = \mu_B \sqrt{l(l+1)} \,, \tag{7.11}$$

where $\mu_{\rm B}$ is the **Bohr magneton** defined by:

$$\mu_{\rm B} = \frac{e\hbar}{2m_{\rm e}} = 9.27 \times 10^{-24}\,{\rm JT}^{-1}. \tag{7.12}$$

This shows that the size of atomic dipoles is of order $\mu_{\rm B}$. In many cases we are interested in the z component of the magnetic dipole, which is given from eqns 7.2 and 7.4 as:

$$\mu_z = -\frac{e}{2m_{\rm e}} L_z = -\mu_{\rm B} m_l\,, \tag{7.13}$$

where m_l is the orbital magnetic quantum number.

7.2 Spin magnetism

We have seen in Section 5.2.2 that electrons also have spin angular momentum. The deflections measured in the Stern-Gerlach experiment (see Fig. 5.1) enabled the magnitude of the magnetic moment due to the spin angular momentum to be determined. The component along the z axis was found to obey:

$$\mu_z = -g_{\rm s}\mu_{\rm B} m_s\,, \tag{7.14}$$

where $g_{\rm s}$ is the **g-value** of the electron, and $m_s = \pm 1/2$ is the magnetic quantum number due to spin. This is identical in form to eqn 7.13 apart from the factor of $g_{\rm s}$. The experimental value of $g_{\rm s}$ was found to be close to 2. The Dirac equation predicts that $g_{\rm s}$ should be exactly equal to 2, and more recent calculations based on quantum electrodynamics (QED) give a value of $2.0023192\cdots$, which agrees very accurately with the most precise experimental data.

It should be noted that other branches of physics sometimes use a different sign convention in which the electron spin g-factor is negative. The negative charge of the electron is factored into the g-factor, which is defined by:

$$\boldsymbol{\mu}_{\rm spin} = g_{\rm e}\frac{\mu_{\rm B}}{\hbar}\boldsymbol{s}\,, \tag{7.15}$$

where $\boldsymbol{s}$ is the spin angular momentum, and $\mu_{\rm B}/\hbar = e/2m_{\rm e}$ is the magnitude of the electron gyromagnetic ratio. This implies:

$$\mu_z = g_{\rm e}\frac{\mu_{\rm B}}{\hbar} s_z = g_{\rm e}\mu_{\rm B} m_s\,. \tag{7.16}$$

On comparing to eqn 7.14, it is apparent that $g_{\rm s}$ and $g_{\rm e}$ are related to each other through:

$$g_{\rm s} = |g_{\rm e}| = -g_{\rm e}\,. \tag{7.17}$$

The convention in which the sign of the g-factor relates to the charge of the particle is frequently used in tables of fundamental constants. However, in atomic physics we are almost always dealing with electrons, and so it is more convenient to use the the positive value $g_{\rm s}$ rather than the negative one $g_{\rm e}$.

7.3 Spin-orbit coupling

The fact that electrons in atoms have both orbital and spin angular momentum leads to a new interaction term in the Hamiltonian called **spin-orbit coupling**. Sophisticated theories of spin-orbit coupling (e.g. those based on the Dirac equation) indicate that it is actually a relativistic effect. At this stage it is more useful to consider spin-orbit coupling as the interaction between the magnetic field due to the orbital motion of the electron and the magnetic moment due to its spin. This more intuitive approach is the one we adopt here. We start by giving a simple order of magnitude estimate based on the semi-classical Bohr model, and then take a more general approach that works for the fully quantum mechanical picture.

7.3.1 Spin-orbit coupling in the Bohr model

The easiest way to understand the spin-orbit coupling is to consider the single electron of a hydrogen atom in a Bohr-like circular orbit around the nucleus, and then shift the origin to the electron, as indicated in Fig. 7.2. In this frame, the electron is stationary and the nucleus is moving in a circular orbit of radius r_n. The orbit of the nucleus is equivalent to a current loop, which produces a magnetic field at the origin. Now the magnetic field produced by a circular loop of radius r carrying a current i is given by:

$$B_z = \frac{\mu_0 i}{2r}\,, \tag{7.18}$$

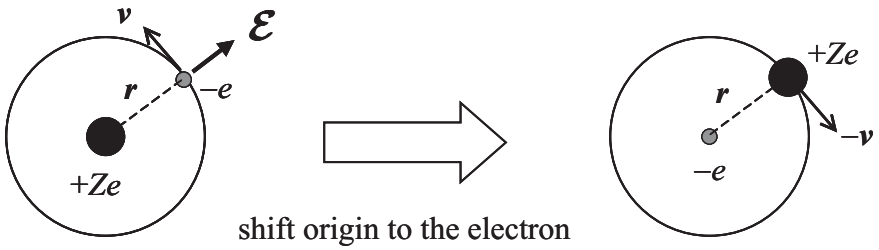

Figure 7.2: An electron moving with velocity $\boldsymbol{v}$ through the electric field $\boldsymbol{\mathcal{E}}$ of the nucleus experiences a magnetic field equal to $(\boldsymbol{\mathcal{E}}\times\boldsymbol{v})/c^2$. The magnetic field can be understood by shifting the origin to the electron and calculating the magnetic field due to the orbital motion of the nucleus around the electron. The velocity of the nucleus in this frame is equal to $-\boldsymbol{v}$.

where z is taken to be the direction perpendicular to the loop. As in Section 7.1, the current i is given by the charge Ze divided by the orbital period $T = 2\pi r/v$. On substituting for the velocity and radius in the Bohr model from eqns 2.12 and 2.13, we find:

$$B_z = \frac{\mu_0 Z e v_n}{4\pi r_n^2} = \left(\frac{Z^4}{n^5}\right)\frac{\mu_0 \alpha c e}{4\pi a_0^2}\,, \tag{7.19}$$

where $\alpha = e^2/2\epsilon_0 hc \approx 1/137$ is the **fine structure constant** defined in eqn 2.15. For hydrogen with $Z = n = 1$, this gives $B_z \approx 12\,\text{Tesla}$, which is a large field.

The electron at the origin experiences this orbital field and we thus have a magnetic interaction energy of the form:

$$\Delta E_{\rm so} = -\boldsymbol{\mu}_{\rm spin}\cdot\boldsymbol{B}_{\rm orbital}\,, \tag{7.20}$$

which, from eqn 7.14, becomes:

$$\Delta E_{\rm so} = g_{\rm s}\mu_{\rm B} m_s B_z = \pm\mu_{\rm B} B_z\,, \tag{7.21}$$

where we have used $g_{\rm s} = 2$ and $m_s = \pm 1/2$ in the last equality. By substituting from eqn. 7.19 and making use of eqn 7.12, we find:

$$|\Delta E_{\rm so}| = \left(\frac{Z^4}{n^5}\right)\frac{\mu_0 \alpha c e^2 \hbar}{8\pi m_{\rm e} a_0^2} \equiv \alpha^2 \frac{Z^2}{n^3}|E_n|\,, \tag{7.22}$$

where E_n is the quantized energy given by eqn 2.6. For the $n = 2$ orbit of hydrogen, this gives:[1]

$$|\Delta E_{\rm so}| = \alpha^2 R_H/2^5 = 13.6\,\text{eV}/32\times 137^2 = 0.02\,\text{meV} \equiv 0.2\,\text{cm}^{-1}\,.$$

This shows that the spin-orbit interaction is about 10^5 times smaller than the gross structure energy in hydrogen. Note that the relative size of the spin-orbit interaction grows as Z^2, so that spin-orbit effects are expected to become more important in heavier atoms, which is indeed the case.

A connection with relativistic theories can be made by noting that eqn 7.22 can be re-written using eqn 2.13 as

$$|\Delta E_{\rm so}| = \left(\frac{v_n}{c}\right)^2 \frac{|E_n|}{n}\,. \tag{7.23}$$

This shows that the spin-orbit interaction energy depends on v^2/c^2, just as we would expect for a relativistic correction to the Bohr model. This is hardly surprising, given that Dirac tells us that we should really think of spin-orbit coupling as a relativistic effect.

7.3.2 Spin-orbit coupling beyond the Bohr model

In this sub-section we repeat the calculation above but without making use of the semi-classical results from the Bohr model. The electrons experience a magnetic field as they move through the electric field of the nucleus. If the electron velocity is $\boldsymbol{v}$, it will see the nucleus orbiting around it with a velocity of $-\boldsymbol{v}$,

[1] The actual spin-orbit splitting of the hydrogen $n = 2$ level is shown in Fig. 7.4, and is about twice the size calculated from the Bohr model. The equivalent Bohr-model value for the $n = 1$ orbit is $0.7\,\text{meV}$ ($6\,\text{cm}^{-1}$). However, this is not very meaningful, as the $n = 1$ level only has $l = 0$, and so the spin-orbit interaction is, in fact, zero.

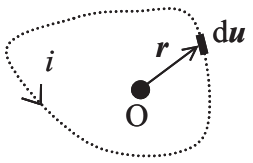

Figure 7.3: The magnetic field at the origin O due to a loop carrying a current i is calculated by the Biot-Savart law given in Eq. 7.24. The field points out of the paper.

as shown in Fig. 7.2. The magnetic field generated at the electron can be calculated by the Biot-Savart law as shown by Fig. 7.3. This gives the magnetic field at the origin of a loop carrying a current i as:

$$\boldsymbol{B} = \frac{\mu_0}{4\pi} \oint_{\text{loop}} i \, \frac{\mathrm{d}\boldsymbol{u} \times \boldsymbol{r}}{r^3} \,, \tag{7.24}$$

where $\mathrm{d}\boldsymbol{u}$ is an orbital path element. For simplicity we consider the case of a circular orbit with constant r. In this case we have:

$$\oint i \, \mathrm{d}\boldsymbol{u} = \oint \frac{\mathrm{d}q}{\mathrm{d}t} \, \mathrm{d}\boldsymbol{u} = Ze \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = Ze(-\boldsymbol{v}) \,.$$

We thus obtain:

$$\boldsymbol{B} = -\frac{\mu_0}{4\pi} \frac{Ze}{r^3} \boldsymbol{v} \times \boldsymbol{r} = \frac{\mu_0}{4\pi} \frac{Ze}{r^3} \boldsymbol{r} \times \boldsymbol{v} \,. \tag{7.25}$$

For a Coulomb field the electric field $\boldsymbol{\mathcal{E}}$ is given by:

$$\boldsymbol{\mathcal{E}} = \frac{Ze}{4\pi\epsilon_0 r^2} \hat{\boldsymbol{r}} = \frac{Ze}{4\pi\epsilon_0 r^3} \boldsymbol{r} \,, \tag{7.26}$$

where the hat symbol on $\hat{\boldsymbol{r}}$ in the first equality indicates that it is a unit vector. On combining equations 7.25 and 7.26 we obtain:

$$\boldsymbol{B} = \mu_0 \epsilon_0 \, \boldsymbol{\mathcal{E}} \times \boldsymbol{v} \,. \tag{7.27}$$

We know from Maxwell's equations that $\mu_0\epsilon_0 = 1/c^2$, and so we can re-write this as:

$$\boldsymbol{B} = \frac{1}{c^2} \, \boldsymbol{\mathcal{E}} \times \boldsymbol{v} \,. \tag{7.28}$$

The same formula can also be derived for the more general case of non-circular orbits and non-Coulombic electric fields such as those found in multi-electron atoms.

The spin-orbit interaction energy is given by:

$$\Delta E_{\text{so}} = -\boldsymbol{\mu}_{\text{spin}} \cdot \boldsymbol{B}_{\text{orbital}} \,, \tag{7.29}$$

where $\boldsymbol{\mu}_{\text{spin}}$ is the magnetic moment due to spin, which is given by:

$$\boldsymbol{\mu}_{\text{spin}} = -g_{\text{s}} \frac{|e|}{2m_{\text{e}}} \boldsymbol{s} = -g_{\text{s}} \frac{\mu_{\text{B}}}{\hbar} \boldsymbol{s} \,. \tag{7.30}$$

On substituting Eqs. 7.28 and 7.30 into Eq. 7.29, we obtain:

$$\Delta E_{\text{so}} = \frac{g_{\text{s}} \mu_{\text{B}}}{\hbar c^2} \boldsymbol{s} \cdot (\boldsymbol{\mathcal{E}} \times \boldsymbol{v}) \,. \tag{7.31}$$

If we have a **central field** (ie the potential V is a function of r only), we can write:[2]

$$\boldsymbol{\mathcal{E}} = \frac{1}{e} \frac{\boldsymbol{r}}{r} \frac{\mathrm{d}V}{\mathrm{d}r} \,. \tag{7.32}$$

On making use of this, the spin-orbit energy becomes:

$$\Delta E_{\text{so}} = \frac{g_{\text{s}} \mu_{\text{B}}}{\hbar c^2 e m_{\text{e}}} \left(\frac{1}{r} \frac{\mathrm{d}V}{\mathrm{d}r} \right) \boldsymbol{s} \cdot (\boldsymbol{r} \times \boldsymbol{p}) \,, \tag{7.33}$$

[2]It is easy to verify that this works for a Coulomb field where $V = -Ze^2/4\pi\epsilon_0 r$ and $\boldsymbol{\mathcal{E}}$ is given by eqn 7.26.

where we have substituted $\boldsymbol{v} = \boldsymbol{p}/m_{\rm e}$. On recalling that the angular momentum $\boldsymbol{l}$ is defined as $\boldsymbol{r} \times \boldsymbol{p}$, we find:

$$\Delta E_{\rm so} = \frac{g_{\rm s}\mu_{\rm B}}{\hbar c^2 e m_{\rm e}} \left(\frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right) \boldsymbol{s}\cdot\boldsymbol{l}\,. \tag{7.34}$$

This calculation of $\Delta E_{\rm so}$ does not take proper account of relativistic effects. In particular, we moved the origin from the nucleus to the electron, which is not really allowed because the electron is accelerating all the time and is therefore not an inertial frame. The translation to a rotating frame gives rise to an extra effect called the **Thomas precession** which reduces the energy by a factor of 2. (See Eisberg and Resnick, Appendix O.) On taking the Thomas precession into account, and recalling that $\mu_{\rm B} = e\hbar/2m_{\rm e}$, we obtain the final result:

$$\Delta E_{\rm so} = \frac{g_{\rm s}}{2}\frac{1}{2c^2 m_{\rm e}^2} \left(\frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right) \boldsymbol{l}\cdot\boldsymbol{s}\,. \tag{7.35}$$

This is the same as the result derived from the Dirac equation, except that $g_{\rm s}$ is exactly equal to 2 in Dirac's theory. Equation 7.35 shows that the spin and orbital angular momenta are coupled together. If we have a simple Coulomb field and take $g_{\rm s} = 2$, we find

$$\Delta E_{\rm so} = \frac{Ze^2}{8\pi\epsilon_0 c^2 m_{\rm e}^2} \left(\frac{1}{r^3} \right) \boldsymbol{l}\cdot\boldsymbol{s}\,. \tag{7.36}$$

We can use this formula for hydrogenic atoms, while we can use the more general form given in Eq. 7.35 for more complicated multi-electron atoms where the potential will differ from the Coulombic $1/r$ dependence due to the repulsion between the electrons.

7.4 Evaluation of the spin-orbit energy for hydrogen

The magnitude of the spin-orbit energy can be calculated from eqn 7.35 as:

$$\Delta E_{\rm so} = \frac{1}{2c^2 m_{\rm e}^2} \left\langle \frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right\rangle \langle \boldsymbol{l}\cdot\boldsymbol{s} \rangle\,, \tag{7.37}$$

where we have taken $g_{\rm s} = 2$, and the $\langle \cdots \rangle$ notation indicates that we take expectation values:

$$\left\langle \frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right\rangle = \iiint \psi^*_{nlm} \left(\frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right) \psi_{nlm}\, r^2 \sin\theta\, {\rm d}r{\rm d}\theta{\rm d}\phi\,. \tag{7.38}$$

The function $({\rm d}V/{\rm d}r)/r$ depends only on r, and so we are left to calculate an integral over r only:

$$\left\langle \frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right\rangle = \int_0^\infty |R_{nl}(r)|^2 \left(\frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right) r^2 {\rm d}r\,, \tag{7.39}$$

where $R_{nl}(r)$ is the radial wave function. This integral can be evaluated exactly for the case of the Coulomb field in hydrogen where $({\rm d}V/{\rm d}r)/r \propto 1/r^3$, and the radial wave functions are known exactly. (See Table 2.3.) We then have, for $l \geq 1$:

$$\left\langle \frac{1}{r}\frac{{\rm d}V}{{\rm d}r} \right\rangle \propto \left\langle \frac{1}{r^3} \right\rangle = \frac{Z^3}{a_0^3 n^3 l(l+\frac{1}{2})(l+1)}\,. \tag{7.40}$$

This shows that we can re-write eqn 7.37 in the form:

$$\Delta E_{\rm so} = C_{nl}\, \langle \boldsymbol{l}\cdot\boldsymbol{s} \rangle\,, \tag{7.41}$$

where C_{nl} is a constant that depends only on n and l.

We can evaluate $\langle \boldsymbol{l}\cdot\boldsymbol{s} \rangle$ by realizing from eqn 5.22 that we must have:

$$\boldsymbol{j}^2 = (\boldsymbol{l}+\boldsymbol{s})^2 = \boldsymbol{l}^2 + \boldsymbol{s}^2 + 2\boldsymbol{l}\cdot\boldsymbol{s}\,. \tag{7.42}$$

This implies that:

$$\langle \boldsymbol{l}\cdot\boldsymbol{s} \rangle = \left\langle \frac{1}{2}(\boldsymbol{j}^2 - \boldsymbol{l}^2 - \boldsymbol{s}^2) \right\rangle = \frac{\hbar^2}{2}[j(j+1) - l(l+1) - s(s+1)]\,. \tag{7.43}$$

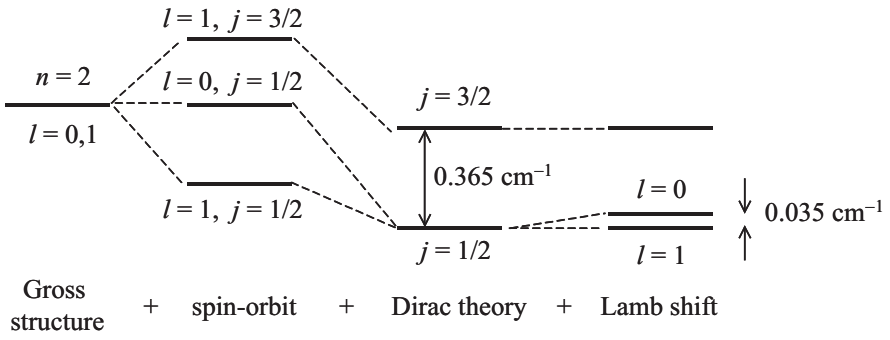

Figure 7.4: Fine structure in the $n = 2$ level of hydrogen.

We therefore find:

$$\Delta E_{\rm so} = C'_{nl}\,[j(j+1) - l(l+1) - s(s+1)]\,, \tag{7.44}$$

where $C'_{nl} = C_{nl}\hbar^2/2$. On using eqn 7.40 we obtain the final result for states with $l \geq 1$:

$$\Delta E_{\rm so} = -\frac{\alpha^2 Z^2}{2n^2} E_n \frac{n}{l(l+\frac{1}{2})(l+1)}[j(j+1) - l(l+1) - s(s+1)]\,, \tag{7.45}$$

where $\alpha \approx 1/137$ is the **fine structure constant**, and $E_n = -R_{\rm H}Z^2/n^2$ is equal to the gross energy. For states with $l = 0$ it is apparent from eqn 7.37 that $\Delta E_{\rm so} = 0$.

The fact that j takes values of $l+1/2$ and $l-1/2$ for $l \geq 1$ means that the spin-orbit interaction splits the two j states with the same value of l. We thus expect the electronic states of hydrogen with $l \geq 1$ to split into doublets. However, the actual fine structure of hydrogen is more complicated for two reasons:

1. States with the same n but different l are degenerate.

2. The spin-orbit interaction is small.

The first point is a general property of pure one-electron systems, and the second follows from the scaling of $\Delta E_{\rm so}/E_n$ with Z^2. A consequence of point 2 is that other relativistic effects that have been neglected up until now are of a similar magnitude to the spin-orbit coupling. In atoms with higher values of Z, the spin-orbit coupling is the dominant relativistic correction, and we can neglect the other effects.

The fine structure of the $n = 2$ level in hydrogen is illustrated in figure 7.4. The fully relativistic Dirac theory predicts that states with the same j are degenerate. The degeneracy of the two $j = 1/2$ states is ultimately lifted by a quantum electrodynamic (QED) effect called the Lamb shift. The complications of the fine structure of hydrogen due to other relativistic and QED effects means that hydrogen is not the paradigm for understanding spin-orbit effects. The alkali metals considered below are in fact simpler to understand.

7.5 Spin-orbit coupling in alkali atoms

Alkali atoms have a single valence electron outside close shells. Closed shells have no angular momentum, and so the angular momentum state $|L, S, J\rangle$ of the atom is determined entirely by the valence electron. By analogy with the results for hydrogen given in eqns 7.41 and eqn 7.44, we can write the spin-orbit interaction term as:

$$\Delta E_{SO} \propto \langle \boldsymbol{L} \cdot \boldsymbol{S} \rangle \propto [J(J+1) - L(L+1) - S(S+1)]\,. \tag{7.46}$$

It follows immediately that the spin-orbit energy is zero when the valence electron is in an s-shell, since $\boldsymbol{L} \cdot \boldsymbol{S} = 0$ when $L = 0$. (Alternatively: $J = S$ if $L = 0$, so $J(J+1) - L(L+1) - S(S+1) = 0$.)

Now consider the case when the valence electron is in a shell with $l \neq 0$. We now have $L = l$ and $S = 1/2$, so that $\boldsymbol{L} \cdot \boldsymbol{S} \neq 0$. J has two possible values, namely $J = L \oplus S = L \oplus 1/2 = L \pm 1/2$. On writing eqn 7.46 in the form:

$$\Delta E_{SO} = C\,[J(J+1) - L(L+1) - S(S+1)]\,, \tag{7.47}$$

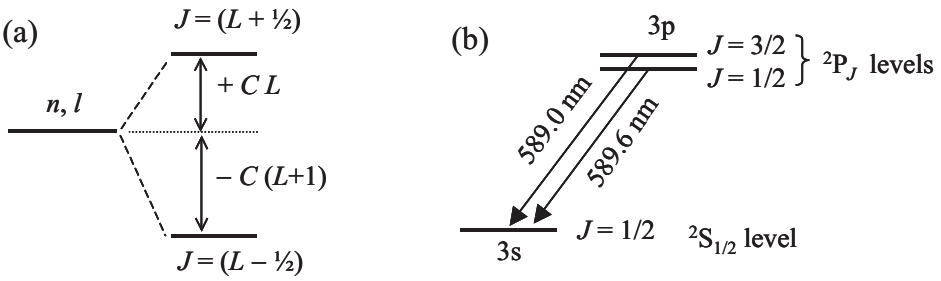

Figure 7.5: Spin-orbit interactions in alkali atoms. (a) The spin-orbit interaction splits the nl states into a doublet if $l \neq 0$. (b) Fine structure in the yellow sodium D lines.

the spin-orbit energy of the $J = (L + 1/2)$ state is given by:

$$\Delta E_{\rm so} = C\,\left[(L+\frac{1}{2})(L+\frac{3}{2}) - L(L+1) - \frac{1}{2}\cdot\frac{3}{2}\right] = +CL\,,$$

while for the $J = (L - 1/2)$ level we have:

$$\Delta E_{\rm so} = C\,\left[(L-\frac{1}{2})(L+\frac{1}{2}) - L(L+1) - \frac{1}{2}\cdot\frac{3}{2}\right] = -C(L+1)\,.$$

Hence the term defined by the quantum numbers n and l is split by the spin-orbit coupling into two new states, as illustrated in figure 7.5(a). This gives rise to the appearance of doublets in the atomic spectra. The magnitude of the splitting is smaller than the gross energy by a factor $\sim \alpha^2 = 1/137^2$. (See Eq. 7.45.) This is why these effects are called "fine structure", and α is called the "fine structure constant".

As an example, let us consider sodium, which has 11 electrons, with one valence electron outside filled 1s, 2s and 2p shells. It can therefore be treated as a one electron system, provided we remember that this is only an approximation. One immediate consequence is that the differing l states arising from the same value of n are not degenerate as they are in hydrogen. (See section 4.5.) The bright yellow D lines of sodium correspond to the 3p $\rightarrow$ 3s transition.[3]

It is well known that the D-lines actually consist of a doublet, as shown in Fig. 7.5(b). The doublet arises from the spin-orbit coupling. The ground state is a $^2{\rm S}_{1/2}$ level with zero spin-orbit splitting. The excited state is split into the two levels derived from the different J values for $L = 1$ and $S = 1/2$, namely the $^2{\rm P}_{3/2}$ and $^2{\rm P}_{1/2}$ levels. The two transitions in the doublet are therefore:

$$^2{\rm P}_{3/2} \rightarrow {}^2{\rm S}_{1/2}$$

and

$$^2{\rm P}_{1/2} \rightarrow {}^2{\rm S}_{1/2}\,.$$

The energy difference of $17\,{\rm cm}^{-1}$ between them arises from the spin-orbit splitting of the two J states of the ^{2}P term.

Similar arguments can be applied to the other alkali elements. The spin–orbit energy splittings of their first excited states are tabulated in Table 7.1. Note that the splitting increases with Z, and that the splitting energy is roughly proportional to Z^2, as shown in Fig. 7.6. This is an example of the fact that spin–orbit interactions generally increase with the atomic number, so that the spin–orbit coupling is stronger in heavier elements.

It should be pointed out that the ordering of the levels shown in Fig. 7.5(a) assumes that the constant C in eqn 7.47 is positive, so that the level with $J = L + 1/2$ lies above the one with $J = L - 1/2$. This is true in most cases, but there are some exceptions. For example, C is negative for the 3d states of sodium, so that the $^2{\rm D}_{5/2}$ level lies *below* $^2{\rm D}_{3/2}$. The 4d term of potassium is also inverted. There is no simple reason why this should be so. It depends on complicated exchange effects.

[3]The first optical transition of an alkali from its ground state is called a "D line". This terminology dates back to Fraunhofer's classification of the absorption lines observed in the Sun's spectrum. It originally applied only to sodium, but is now used for all alkalis: see Table 7.1.

Element	Z	Ground state	1st excited state	Transition	ΔE (cm^{-1})
Lithium	3	[He] 2s	2p	2p →2s	0.33
Sodium	11	[Ne] 3s	3p	3p →3s	17
Potassium	19	[Ar] 4s	4p	4p →4s	58
Rubidium	37	[Kr] 5s	5p	5p →5s	238
Cesium	55	[Xe] 6s	6p	6p →6s	554

Table 7.1: Spin-orbit splitting ΔE of the D lines of the alkali elements. The energy splitting is equal to the difference of the energies of the $J = 3/2$ and $J = 1/2$ levels of the first excited state.

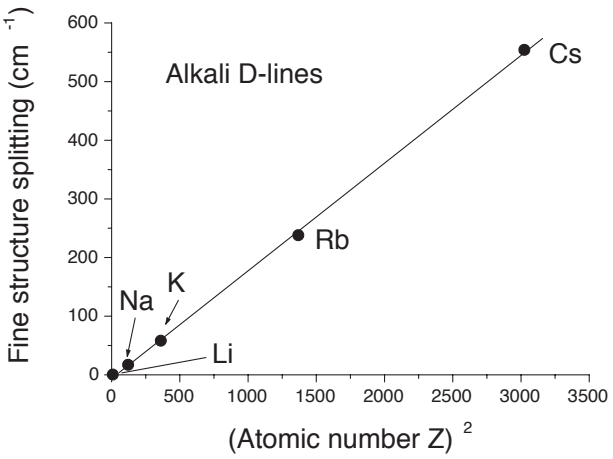

Figure 7.6: Spin-orbit splitting of the first excited state of the alkali atoms versus Z^2, as determined by the fine structure splitting of the D-lines. (See Table 7.1.)

7.6 Spin-orbit coupling in many-electron atoms

We have seen in Chapter 5 that atoms with more than one valence electron can have different types of angular momentum coupling. We restrict our attention here to atoms with LS-coupling, which is the most common type, as explained in Section 5.7. In LS-coupling, the residual electrostatic interaction couples the orbital and spin angular momenta together according to eqns 5.29 and 5.30. The resultants are then coupled together to give the total angular momentum $\boldsymbol{J}$ according to

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}\,. \tag{7.48}$$

The rules for coupling of angular momenta produce several J states for each LS-term, with J running from $L + S$ down to $|L - S|$ in integer steps.[4] These J states experience different spin-orbit interactions, and so are shifted in energy from each other. Hence the spin-orbit coupling splits the J states of a particular LS-term into fine structure multiplets.

The splitting of the J states can be evaluated as follows. The spin-orbit interaction takes the form:

$$\Delta E_{\rm so} = -\boldsymbol{\mu}_{\rm spin} \cdot \boldsymbol{B}_{\rm orbital} \propto \langle \boldsymbol{L} \cdot \boldsymbol{S} \rangle\,, \tag{7.49}$$

which implies (cf. eqns 7.41 – 7.44):

$$\Delta E_{SO} = C_{LS}\,[J(J+1) - L(L+1) - S(S+1)]\,. \tag{7.50}$$

It follows from eqn 7.50 that levels with the same L and S but different J are separated by an energy which is proportional to J. This is called the **interval rule**. Figure 5.3 shows an example of the interval rule for the ^{3}P term of the (3s,3p) configuration of magnesium.

[4] There is only one J state, and hence no fine structure splitting, when one or both of L or S are zero.

7.7 Nuclear effects in atoms

For most of the time in atomic physics we just take the nucleus to be a heavy charged particle sitting at the centre of the atom. However, careful analysis of the spectral lines can reveal small effects that give us direct information about the nucleus. The main effects that can be observed generally fall into two categories, namely **isotope shifts** and **hyperfine structure**.

7.7.1 Isotope shifts

There are two main processes that give rise to isotope shifts in atoms, namely mass effects and field effects.

Mass effects The mass m that enters the Schrödinger equation is the *reduced* mass, not the bare electron mass m_e (cf. eqn 2.5). Changes in the nuclear mass therefore make small changes to m and hence to the atomic energies.

Field effects Electrons in s shells have a finite probability of penetrating the nucleus, and are therefore sensitive to its charge distribution.

Both effects cause small shifts in the wavelengths of the spectral lines from different isotopes of the same element. The heavy isotope of hydrogen, namely deuterium, was discovered in this way through its mass effect.

7.7.2 Hyperfine structure

In high–resolution spectroscopy, it is necessary to consider effects relating to the magnetic interaction between the electron angular momentum ($\boldsymbol{J}$) and the nuclear spin ($\boldsymbol{I}$). This is called the **hyperfine** interaction and arises from the interaction between the magnetic dipole due to the nuclear spin and the magnetic field produced at the nucleus by the electrons.

The spin of the nucleus gives it a magnetic dipole moment which is proportional to $\boldsymbol{I}$:

$$\boldsymbol{\mu}_\mathrm{nucleus} = \gamma_I \boldsymbol{I} = g_I \frac{\mu_\mathrm{N}}{\hbar} \boldsymbol{I}\,, \tag{7.51}$$

where γ_I is the nuclear gyromagnetic ratio, g_I is the nuclear g-factor, and $\mu_\mathrm{N} \equiv e\hbar/2m_\mathrm{P}$ is the nuclear Bohr magneton, with m_P being the proton mass. The value of μ_N in S.I. units is 5.050783×10^{-27} A m^2. There are several interesting points that can be made here in comparison to the equivalent result for electrons:

- The nuclear gyromagnetic ratio is positive, whereas the electron gyromagnetic ratio is negative (see eqn 7.4). This follows from the positive charge of the nucleus, and means that nuclear dipoles are parallel to the nuclear spin. This contrasts with electrons, where the dipole is anti-parallel to the angular momentum on account of the negative sign of the electron charge.
- The nuclear gyromagnetic ratio is about 2000 times smaller than the electron gyromagnetic ratio on account of the heavier proton mass.
- The presence of the nuclear g-factor in eqn 7.51 highlights the quantum-mechanical origin of nuclear spin. Following the convention in which the sign of the g-factor relates to the sign of the particle's charge as discussed in Section 7.2, then the g-factors of protons and neutrons are $+5.5857$ and -3.8261 respectively. These non-integer values point to the fact that protons and neutrons are actually composite rather than elementary particles. The negative, non-zero value for the neutron is particularly striking, given that the neutron is uncharged.

The hyperfine interaction is of the form:

$$\Delta E_\mathrm{hyperfine} = -\boldsymbol{\mu}_\mathrm{nucleus} \cdot \boldsymbol{B}_\mathrm{electron} \propto \langle \boldsymbol{I} \cdot \boldsymbol{J} \rangle\,. \tag{7.52}$$

The presence of $\boldsymbol{J}$ in this interaction term arises from the fact that the magnetic field generated by the electrons at the nucleus depends on the total electronic angular momentum. The magnitude of the hyperfine interaction is very small because the nuclear dipole is about 2000 times smaller than that of the electron. The splittings are therefore about three orders of magnitude smaller than the fine structure splittings: hence the name "hyperfine".

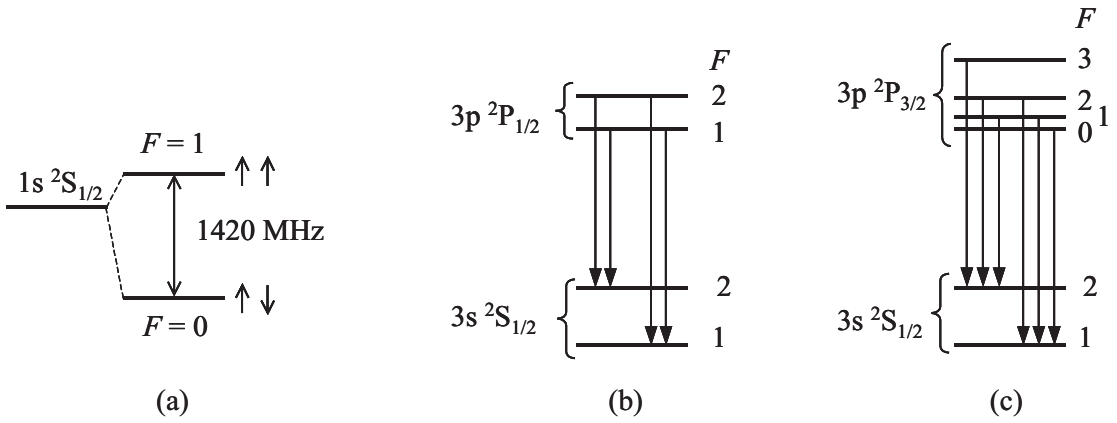

Figure 7.7: (a) Hyperfine structure of the 1s ground state of hydrogen. The arrows indicate the relative directions of the electron and nuclear spin. (b) Hyperfine transitions for the sodium D_1 line. (c) Hyperfine transitions for the sodium D_2 line. Note that the hyperfine splittings are not drawn to scale. The splittings of the sodium levels are as follows: $^2S_{1/2}$, 1772 MHz; $^2P_{1/2}$, 190 MHz; $^2P_{3/2}$ $(3 \to 2)$, 59 MHz; $^2P_{3/2}$ $(2 \to 1)$, 34 MHz; $^2P_{3/2}$ $(1 \to 0)$, 16 MHz.

Hyperfine states are labelled by the total angular momentum $\boldsymbol{F}$ of the whole atom (i.e. nucleus plus electrons), where

$$\boldsymbol{F} = \boldsymbol{I} + \boldsymbol{J}. \tag{7.53}$$

This implies that the hyperfine interaction is of the form (cf. eqn 7.43):

$$\Delta E_{\text{hyperfine}} = \frac{A}{2}\left(F(F+1) - I(I+1) - J(J+1)\right) , \tag{7.54}$$

where A is the hyperfine constant. In analogy with the $|LSJ\rangle$ states of fine structure, the electric dipole selection rule for transitions between hyperfine states is:

$$\Delta F = 0, \pm 1 , \tag{7.55}$$

with the exception that $F = 0 \to 0$ transitions are forbidden. Let us consider two examples to see how this works.

The hydrogen 21 cm line

Consider the ground state of hydrogen. The nucleus consists of just a single proton, and we therefore have $I = 1/2$. The hydrogen ground state is the 1s $^2S_{1/2}$ term, which has $J = 1/2$. The hyperfine quantum number F is then found from $F = I \oplus J = 1/2 \oplus 1/2 = 1$ or 0. These two hyperfine states correspond to the cases in which the spins of the electron and the nucleus are aligned parallel ($F = 1$) or antiparallel ($F = 0$). The two F states are split by the hyperfine interaction by 0.0475 cm^{-1} (5.9×10^{-6} eV). (See Fig. 7.7(a).) Transitions between these levels occur at 1420 MHz ($\lambda = 21$ cm), and are very important in radio astronomy. Radio frequency transitions such as these are also routinely exploited in **nuclear magnetic resonance** (NMR) spectroscopy. (See section 8.3.)

Hyperfine structure of the sodium D lines

The sodium D lines originate from 3p $\to$ 3s transitions. As discussed in Section 7.5, there are two lines with energies split by the spin-orbit coupling, as indicated in Fig. 7.5(b).

Consider first the lower energy D_1 line, which is the $^2P_{1/2} \to ^2S_{1/2}$ transition. The nucleus of sodium has $I = 3/2$, and so we have $F = 3/2 \oplus 1/2 = 2$ or 1 for both the upper and lower levels of the transition, as shown in Fig. 7.7(b). Note that the hyperfine splittings are not drawn to scale in Fig. 7.7(b): the splitting of the $^2S_{1/2}$ level is 1772 MHz, which is much larger than that of the $^2P_{1/2}$, namely 190 MHz. This is a consequence of the fact that s-electrons have higher probability densities at the nucleus, and hence experience stronger hyperfine interactions. All four transitions are allowed by the selection rules, and so we observe four lines. Since the splitting of the upper and lower levels are so different, we obtain

two doublets with relative frequencies of (0, 190) MHz and (1772, 1962) MHz. These splittings should be compared to the much larger ($\sim 5 \times 10^{11}$ Hz) splitting between the two J states caused by the spin-orbit interaction. Since the hyperfine splittings are much smaller, they are not routinely observed in optical spectroscopy, and specialized techniques using narrow band lasers are typically employed nowadays.

Now consider the higher energy D_2 line, which is the $^2P_{3/2} \rightarrow ^2 S_{1/2}$ transition. In the upper level we have $J = 3/2$, and hence $F = I \oplus J = 3/2 \oplus 3/2 = 3$, 2, 1, or 0. There are therefore four hyperfine states for the $^2P_{3/2}$ level, as shown in Fig. 7.7(c). The hyperfine splittings of the $^2P_{3/2}$ level are again much smaller than that of the $^2S_{1/2}$ level, on account of the low probability density of p-electrons near the nucleus. Six transitions are allowed by the selection rules, with the $F = 3 \rightarrow 1$ and $F = 0 \rightarrow 2$ transitions being forbidden by the $|\Delta F| \leq 1$ selection rule. We thus have six hyperfine lines, which split into two triplets at relative frequencies of (0, 34, 59) MHz and (1756, 1772, 1806) MHz.

Reading

Bransden and Joachain, *Physics of Atoms and Molecules*, chapter 5, §9.4–6
Demtröder, *Atoms, Molecules and Photons*, §5.4–8.
Haken and Wolf, *The physics of atoms and quanta*, chapters 12, 20.
Hertel and Schulz, *Atoms, Molecules and Optical Physics, 1*, §6.2 – 6.4, 9.1 – 9.2
Eisberg and Resnick, *Quantum Physics*, chapters 8, 10.
Foot, *Atomic physics*, §2.3, 4.5–6, chapter 6.
Beisser, *Concepts of Modern Physics*, §7.8.

Chapter 8

External fields: the Zeeman and Stark effects

In the previous chapter, we considered the effects of the internal magnetic fields within atoms. We now wish to consider the effects of external fields. Table 8.1 defines the nomenclature of the effects that we shall be considering. We shall start by looking at magnetic fields and then move on to consider electric fields.

Applied field	Field strength	Effect
Magnetic	weak	Zeeman
	strong	Paschen-Back
Electric	all	Stark

Table 8.1: Names of the effects of external fields in atomic physics.

8.1 Magnetic fields

The first person to study the effects of magnetic fields on the optical spectra of atoms was Zeeman in 1896. He observed that the transition lines split when the field is applied. Further work showed that the interaction between the atoms and the field can be classified into two regimes:

- Weak fields: the **Zeeman effect**, either **normal** or **anomalous**;
- Strong fields: the **Paschen-Back effect**.

The "normal" Zeeman effect is so-called because it agrees with the classical theory developed by Lorentz. The "anomalous" Zeeman effect is caused by electron spin, and is therefore a completely quantum result. The criterion for deciding whether a particular field is "weak" or "strong" will be discussed in Section 8.1.3. In practice, we usually work in the weak-field (i.e. Zeeman) limit.

8.1.1 The normal Zeeman effect

The normal Zeeman effect is observed in atoms with no spin. The total spin of an N-electron atom is given by:

$$\boldsymbol{S} = \sum_{i=1}^{N} \boldsymbol{s}_i \,. \tag{8.1}$$

Filled shells have no net spin, and so we only need to consider the valence electrons here. Since all the individual electrons have spin $1/2$, it will not be possible to obtain $S = 0$ from atoms with an odd number

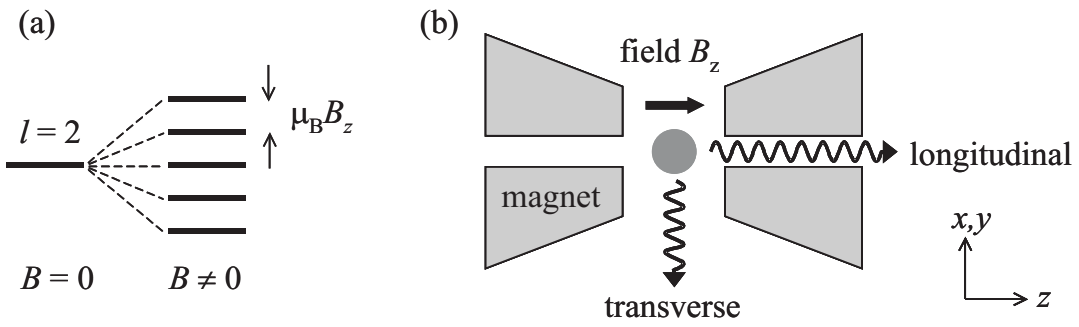

Figure 8.1: The normal Zeeman effect. (a) Splitting of the degenerate m_l states of an atomic level with $l = 2$ by a magnetic field. (b) Definition of longitudinal (Faraday) and transverse (Voigt) observations. The direction of the field defines the z axis.

of valence electrons. However, if there is an even number of valence electrons, we can obtain $S = 0$ states. For example, if we have two valence electrons, then the total spin quantum number $S = 1/2 \oplus 1/2$ can be either 0 or 1. In fact, the ground states of divalent atoms from group II of the periodic table (electronic configuration ns^2) always have $S = 0$ because the two electrons align with their spins antiparallel.

The magnetic moment of an atom with no spin will originate entirely from its orbital motion:

$$\boldsymbol{\mu} = -\frac{\mu_B}{\hbar}\boldsymbol{L}\,, \tag{8.2}$$

where $\mu_B/\hbar = e/2m_e$ is the **gyromagnetic ratio**. (See eqn 7.4.) The interaction energy between a magnetic dipole $\boldsymbol{\mu}$ and a uniform magnetic field $\boldsymbol{B}$ is given by:

$$\Delta E = -\boldsymbol{\mu} \cdot \boldsymbol{B}\,. \tag{8.3}$$

We set up the axes of our spherically-symmetric atom so that the z axis coincides with the direction of the field. In this case we have:

$$\boldsymbol{B} = \begin{pmatrix} 0 \\ 0 \\ B_z \end{pmatrix}\,,$$

and the interaction energy of the atom is therefore:

$$\Delta E = -\mu_z B_z = \mu_B B_z m_l\,, \tag{8.4}$$

where m_l is the orbital magnetic quantum number. Equation 8.4 shows us that the application of an external $\boldsymbol{B}$-field splits the degenerate m_l states evenly. This is why m_l is called the **magnetic quantum number**. The splitting of the m_l states of an $l = 2$ electron is illustrated in Fig. 8.1(a).

The effect of the magnetic field on the spectral lines can be worked out from the splitting of the levels. Consider the transitions between two Zeeman-split atomic levels as shown in Fig. 8.2. The selection rules listed in Table 3.1 of Chapter 3 indicate that we can have transitions with $\Delta m_l = 0$ or ± 1. The gives rise to three transitions whose frequencies are given by:

$$\begin{aligned} h\nu &= h\nu_0 + \mu_B B_z & \Delta m_l &= -1\,, \\ h\nu &= h\nu_0 & \Delta m_l &= 0\,, \\ h\nu &= h\nu_0 - \mu_B B_z & \Delta m_l &= +1\,. \end{aligned} \tag{8.5}$$

This is the same result as that derived by classical theory.

The polarization of the Zeeman lines is determined by the selection rules, and the conditions of observation. If we are looking along the field (**longitudinal** observation), the photons must be propagating in the z direction. (See Fig. 8.1(b).) Light waves are transverse, and so only the x and y polarizations are possible. The z-polarized $\Delta m_l = 0$ line is therefore absent, and we just observe the σ^+ and σ^- circularly polarized $\Delta m_l = \pm 1$ transitions. When observing at right angles to the field (**transverse** observation), all three lines are present. The $\Delta m_l = 0$ transition is linearly polarized parallel to the field, while the $\Delta m_l = \pm 1$ transitions are linearly polarized at right angles to the field. These results are summarized in Table 8.2.[1]

[1] In solid-state physics, the longitudinal and transverse observation conditions are frequently called the **Faraday** and **Voigt** geometries, respectively.

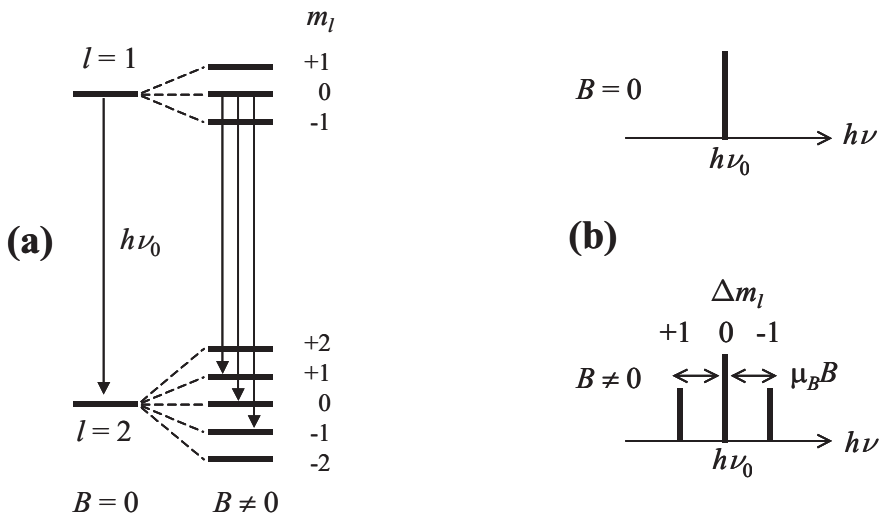

Figure 8.2: The normal Zeeman effect for a p $\rightarrow$ d transition. (a) The field splits the degenerate m_l levels equally. Optical transitions can occur if $\Delta m_l = 0, \pm 1$. (Only the transitions originating from the $m_l = 0$ level of the $l = 1$ state are identified here for the sake of clarity.) (b) The spectral line splits into a triplet when observed transversely to the field. The $\Delta m_l = 0$ transition is unshifted, but the $\Delta m_l = \pm 1$ transitions occur at $(h\nu_0 \mp \mu_{\rm B} B_z)$.

Δm_l	Energy	Polarization	
		Longitudinal observation	Transverse observation
+1	$h\nu_0 - \mu_{\rm B} B$	σ^+	$\boldsymbol{\mathcal{E}} \perp \boldsymbol{B}$
0	$h\nu_0$	not observed	$\boldsymbol{\mathcal{E}} \parallel \boldsymbol{B}$
–1	$h\nu_0 + \mu_{\rm B} B$	σ^-	$\boldsymbol{\mathcal{E}} \perp \boldsymbol{B}$

Table 8.2: The normal Zeeman effect. The last two columns refer to the polarizations observed in longitudinal (Faraday) and transverse (Voigt) observation conditions. The direction of the circular ($\sigma^{\pm}$) polarization in longitudinal observation is defined relative to $\boldsymbol{B}$. In transverse observation, all lines are linearly polarized.

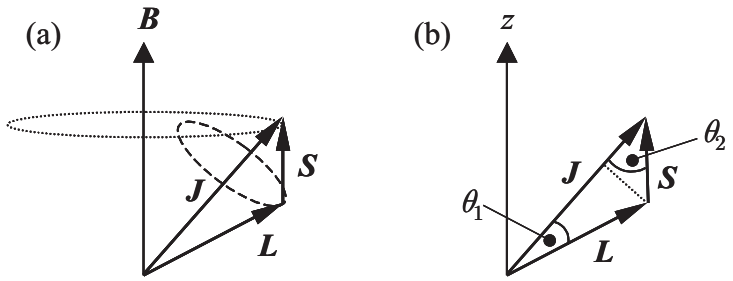

Figure 8.3: (a) Slow precession of $\boldsymbol{J}$ around $\boldsymbol{B}$ in the anomalous Zeeman effect. The spin-orbit interaction causes $\boldsymbol{L}$ and $\boldsymbol{S}$ to precess much more rapidly around $\boldsymbol{J}$. (b) Definition of the projection angles θ_1 and θ_2 used in the calculation of the Landé g factor.

8.1.2 The anomalous Zeeman effect

The anomalous Zeeman effect is observed in atoms with non-zero spin. This will include all atoms with an odd number of electrons. In the LS-coupling regime, the spin-orbit interaction couples the spin and orbital angular momenta together to give the resultant total angular momentum $\boldsymbol{J}$ according to:

$$\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}\,. \tag{8.6}$$

The orbiting electrons in the atom are equivalent to a classical magnetic gyroscope. The torque applied by the field causes the atomic magnetic dipole to precess around $\boldsymbol{B}$, an effect called **Larmor precession**. The external magnetic field therefore causes $\boldsymbol{J}$ to precess slowly about $\boldsymbol{B}$. Meanwhile, $\boldsymbol{L}$ and $\boldsymbol{S}$ precess more rapidly about $\boldsymbol{J}$ due to the spin-orbit interaction. This situation is illustrated in Fig. 8.3(a). The speed of the precession about $\boldsymbol{B}$ is proportional to the field strength. If we turn up the field, the Larmor precession frequency will eventually be faster than the spin-orbit precession of $\boldsymbol{L}$ and $\boldsymbol{S}$ around $\boldsymbol{J}$. This is the point where the behaviour ceases to be Zeeman-like, and we are in the strong field regime of the Paschen-Back effect.

The interaction energy of the atom is equal to the sum of the interactions of the spin and orbital magnetic moments with the field:

$$\Delta E = -\mu_z B_z = -(\mu_z^{\rm spin} + \mu_z^{\rm orbital})B_z = \langle g_{\rm s}\boldsymbol{S}_z + \boldsymbol{L}_z\rangle \frac{\mu_{\rm B}}{\hbar} B_z\,, \tag{8.7}$$

where $g_{\rm s} = 2$, and the symbol $\langle\cdots\rangle$ implies, as usual, that we take expectation values. The normal Zeeman effect is obtained by setting $\boldsymbol{S}_z = 0$ and $\boldsymbol{L}_z = m_l\hbar$ in this formula. In the case of the precessing atomic magnet shown in Fig. 8.3(a), neither $\boldsymbol{S}_z$ nor $\boldsymbol{L}_z$ are constant. Only $\boldsymbol{J}_z = M_J\hbar$ is well-defined. We must therefore first project $\boldsymbol{L}$ and $\boldsymbol{S}$ onto $\boldsymbol{J}$, and then re-project this component onto the z axis. The effective dipole moment of the atom is therefore given by:

$$\boldsymbol{\mu} = -\left\langle |\boldsymbol{L}|\cos\theta_1 \frac{\boldsymbol{J}}{|\boldsymbol{J}|} + 2|\boldsymbol{S}|\cos\theta_2 \frac{\boldsymbol{J}}{|\boldsymbol{J}|}\right\rangle \frac{\mu_{\rm B}}{\hbar}\,, \tag{8.8}$$

where the factor of 2 in the second term comes from the fact that $g_s = 2$. The angles θ_1 and θ_2 that appear here are defined in Fig. 8.3(b), and can be calculated from the scalar products of the respective vectors:

$$\begin{aligned} \boldsymbol{L}\cdot\boldsymbol{J} &= |\boldsymbol{L}|\,|\boldsymbol{J}|\cos\theta_1\,, \\ \boldsymbol{S}\cdot\boldsymbol{J} &= |\boldsymbol{S}|\,|\boldsymbol{J}|\cos\theta_2\,, \end{aligned} \tag{8.9}$$

which implies that:

$$\boldsymbol{\mu} = -\left\langle \frac{\boldsymbol{L}\cdot\boldsymbol{J}}{|\boldsymbol{J}|^2} + 2\frac{\boldsymbol{S}\cdot\boldsymbol{J}}{|\boldsymbol{J}|^2}\right\rangle \frac{\mu_{\rm B}}{\hbar}\,\boldsymbol{J}\,. \tag{8.10}$$

Now equation 8.6 implies that $\boldsymbol{S} = \boldsymbol{J} - \boldsymbol{L}$, and hence that:

$$\boldsymbol{S}\cdot\boldsymbol{S} = (\boldsymbol{J}-\boldsymbol{L})\cdot(\boldsymbol{J}-\boldsymbol{L}) = \boldsymbol{J}\cdot\boldsymbol{J} + \boldsymbol{L}\cdot\boldsymbol{L} - 2\boldsymbol{L}\cdot\boldsymbol{J}\,.$$

We therefore find that:

$$\boldsymbol{L}\cdot\boldsymbol{J} = (\boldsymbol{J}\cdot\boldsymbol{J} + \boldsymbol{L}\cdot\boldsymbol{L} - \boldsymbol{S}\cdot\boldsymbol{S})/2\,,$$

so that:

$$\begin{aligned}\left\langle \frac{\boldsymbol{L}\cdot\boldsymbol{J}}{|\boldsymbol{J}|^2} \right\rangle &= \frac{[J(J+1)+L(L+1)-S(S+1)]\hbar^2/2}{J(J+1)\hbar^2}\,, \\ &= \frac{[J(J+1)+L(L+1)-S(S+1)]}{2J(J+1)}\,. \end{aligned} \tag{8.11}$$

Similarly:

$$\boldsymbol{S}\cdot\boldsymbol{J} = (\boldsymbol{J}\cdot\boldsymbol{J}+\boldsymbol{S}\cdot\boldsymbol{S}-\boldsymbol{L}\cdot\boldsymbol{L})/2\,,$$

and so:

$$\begin{aligned}\left\langle \frac{\boldsymbol{S}\cdot\boldsymbol{J}}{|\boldsymbol{J}|^2} \right\rangle &= \frac{[J(J+1)+S(S+1)-L(L+1)]\hbar^2/2}{J(J+1)\hbar^2}\,, \\ &= \frac{[J(J+1)+S(S+1)-L(L+1)]}{2J(J+1)}\,. \end{aligned} \tag{8.12}$$

We therefore conclude that:

$$\boldsymbol{\mu} = -\left(\frac{[J(J+1)+L(L+1)-S(S+1)]}{2J(J+1)} + 2\frac{[J(J+1)+S(S+1)-L(L+1)]}{2J(J+1)}\right)\frac{\mu_{\rm B}}{\hbar}\,\boldsymbol{J}\,. \tag{8.13}$$

This can be written in the form:

$$\boldsymbol{\mu} = -g_J\frac{\mu_{\rm B}}{\hbar}\boldsymbol{J}\,, \tag{8.14}$$

where g_J is the **Landé g-factor** given by:

$$g_J = 1 + \frac{J(J+1)+S(S+1)-L(L+1)}{2J(J+1)}\,. \tag{8.15}$$

This implies that

$$\mu_z = -g_J\mu_{\rm B}M_J\,, \tag{8.16}$$

and hence that the interaction energy with the field is:

$$\Delta E = -\mu_z B_z = g_J\mu_{\rm B}B_z M_J\,. \tag{8.17}$$

This is the final result for the energy shift of an atomic state in the anomalous Zeeman effect. Note that we just obtain $g_J = 1$ if $S = 0$, as we would expect for an atom with only orbital angular momentum. Similarly, if $L = 0$ so that the atom only has spin angular momentum, we find $g_J = 2$. Classical theories always predict $g_J = 1$. The departure of g_J from unity is caused by the spin part of the magnetic moment, and is a purely quantum effect.

The spectra can be understood by applying the following selection rules on J and M_J:

$$\Delta J = 0, \pm 1\,;$$

$$\Delta M_J = 0, \pm 1\,.$$

These rules have to be applied in addition to the $\Delta l = \pm 1$ and $\Delta S = 0$ rules. (See discussion in § 5.8.)[2] $\Delta J = 0$ transitions are forbidden when $J = 0$ for both states, and $M_J = 0 \to 0$ transitions are forbidden in a $\Delta J = 0$ transition. The transition energy shift is then given by:

$$\begin{aligned} h\Delta\nu &= (h\nu - h\nu_0)\,, \\ &= \left(g_J^{\rm upper}M_J^{\rm upper} - g_J^{\rm lower}M_J^{\rm lower}\right)\mu_{\rm B}B_z\,, \end{aligned} \tag{8.18}$$

where $h\nu_0$ is the transition energy at $B_z = 0$ and the superscripts refer to the upper and lower states respectively.

The polarizations of the transitions follow the same patterns as for the normal Zeeman effect:

- With *longitudinal* observation the $\Delta M_J = 0$ transitions are absent and the $\Delta M_J = \pm 1$ transitions are $\sigma^{\pm}$ circularly polarized.
- With *transverse* observation the $\Delta M_J = 0$ transitions are linearly polarized along the z axis (i.e. parallel to $\boldsymbol{B}$) and the $\Delta M_J = \pm 1$ transitions are linearly polarized in the x-y plane (i.e. perpendicular to $\boldsymbol{B}$).

Level	J	L	S	g_J
$^2P_{3/2}$	3/2	1	1/2	4/3
$^2P_{1/2}$	1/2	1	1/2	2/3
$^2S_{1/2}$	1/2	0	1/2	2

Table 8.3: Landé g-factors evaluated from eqn 8.15 for the levels involved in the sodium D lines.

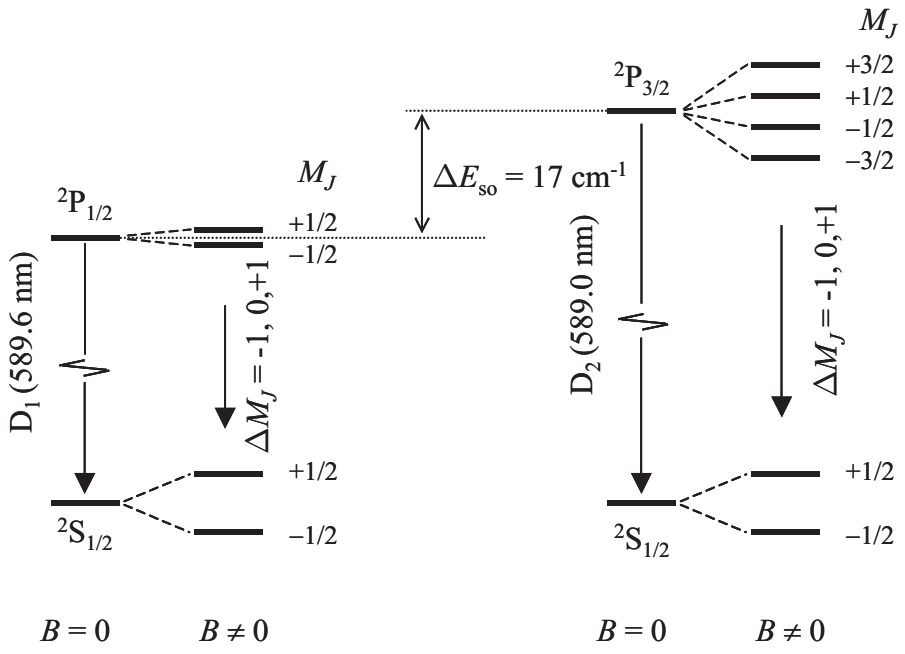

Figure 8.4: Splitting of the sodium D-lines by a weak magnetic field. Note that the Zeeman splittings are smaller than the spin-orbit splitting, as must be the case in the "weak" field limit.

Example: The sodium D lines

The sodium D lines correspond to the 3p $\rightarrow$ 3s transition. At $\boldsymbol{B} = 0$, the spin-orbit interaction splits the upper 3p ^{2}P term into the $^2P_{3/2}$ and $^2P_{1/2}$ levels separated by 17 cm^{-1}. The lower $^2S_{1/2}$ level has no spin-orbit interaction. The Landé g-factors of the levels worked out from eqn 8.15 are given in Table 8.3.

The splitting of the lines in the field is shown schematically in Fig. 8.4. The $^2P_{3/2}$ level splits into four M_J states, while the two $J = 1/2$ levels each split into two states. The splittings are different for each level because of the different Landé factors. On applying the $\Delta M_J = 0, \pm 1$ selection rule, we find four allowed transitions for the D_1 line and six for the D_2. These transitions are listed in Table 8.4.

The results tabulated in Table 8.4 can be compared to those predicted by the normal Zeeman effect. In the normal Zeeman effect we observe three lines with an energy spacing equal to $\mu_B B$. In the anomalous effect, there are more than three lines, and the spacing is different to the classical value: in fact, the lines are not evenly spaced. Furthermore, none of the lines occur at the same frequency as the unperturbed line at $B = 0$.

[2]There are no selection rules on M_L and M_S here because L_z and S_z are not constants of the motion when L and S are coupled by the spin-orbit interaction.

M_J^{upper}	M_J^{lower}	ΔM_J	Transition energy shift	
			D_1 line	D_2 line
$+\frac{3}{2}$	$+\frac{1}{2}$	-1		$+1$
$+\frac{1}{2}$	$+\frac{1}{2}$	0	$-\frac{2}{3}$	$-\frac{1}{3}$
$+\frac{1}{2}$	$-\frac{1}{2}$	-1	$+\frac{4}{3}$	$+\frac{5}{3}$
$-\frac{1}{2}$	$+\frac{1}{2}$	$+1$	$-\frac{4}{3}$	$-\frac{5}{3}$
$-\frac{1}{2}$	$-\frac{1}{2}$	0	$+\frac{2}{3}$	$+\frac{1}{3}$
$-\frac{3}{2}$	$-\frac{1}{2}$	$+1$		-1

Table 8.4: Anomalous Zeeman effect for the sodium D lines. The transition energy shifts are worked out from eqn 8.18 and are quoted in units of $\mu_B B_z$.

8.1.3 The Paschen-Back effect

The Paschen-Back effect is observed at very strong magnetic fields. The criterion for observing the Paschen-Back effect is that the interaction with the external magnetic field should be much stronger than the spin-orbit interaction:

$$\mu_B B_z \gg \Delta E_{so} \,. \tag{8.19}$$

If we satisfy this criterion, then the precession speed around the external field will be much faster than the spin-orbit precession. This means that the interaction with the external field is now the largest perturbation, and so it should be treated first, before the perturbation of the spin-orbit interaction.

Another way to think of the strong-field limit is that it occurs when the external field is much stronger than the internal field of the atom arising from the orbital motion. We saw in Section 7.3 that the internal fields in most atoms are large. For example, the Bohr model predicts an internal field of 12 T for the $n = 1$ shell of hydrogen. (See eqn 7.19.) This is a very strong field, that can only be obtained in the laboratory by using powerful superconducting magnets. This internal field strength is typical of many atoms, and so it will frequently be the case the field required to observe the Paschen-Back effect is so large that we never go beyond the Zeeman regime in the laboratory.[3] For example, in sodium, the field strength equivalent to the spin-orbit interaction for the D-lines is given by:

$$B_z = \frac{\Delta E_{so}}{\mu_B} = \frac{17\,\text{cm}^{-1}}{9.27 \times 10^{-24}\,\text{JT}^{-1}} = 36\,\text{T}\,,$$

which is not achievable in normal laboratory conditions. On the other hand, since the spin-orbit interaction decreases with decreasing atomic number Z, the splitting for the equivalent transition in lithium with $Z = 3$ (i.e. the 2p $\rightarrow$ 2s transition) is only $0.3\,\text{cm}^{-1}$. This means that we can reach the strong field regime for fields $\gg 0.6$ T. This is readily achievable, and allows the Paschen-Back effect to be observed.

In the Paschen-Back effect, the spin-orbit interaction is assumed to be negligibly small, and $\boldsymbol{L}$ and $\boldsymbol{S}$ are therefore no longer coupled together. Each precesses separately around $\boldsymbol{B}$, as sketched in Fig. 8.5. The precession rates for $\boldsymbol{L}$ and $\boldsymbol{S}$ are different because of the different g-values. Hence the magnitude of the resultant $\boldsymbol{J}$ varies with time: the quantum number J is no longer a constant of the motion.

The interaction energy is now calculated by adding the separate contributions of the spin and orbital energies:

$$\Delta E = -\mu_z B_z = -(\mu_z^{\text{orbital}} + \mu_z^{\text{spin}}) B_z = (M_L + g_s M_S)\mu_B B_z \,. \tag{8.20}$$

The shift of the spectral lines is given by:

$$\Delta(h\nu) = (\Delta M_L + g_s \Delta M_S)\mu_B B_z \,. \tag{8.21}$$

[3] There are extremely large magnetic fields present in the Sun due to the circulating plasma currents. This means that the Paschen-Back effect can be observed for elements like sodium in solar spectra.

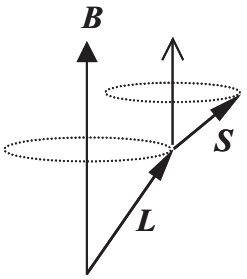

Figure 8.5: Precession of $\boldsymbol{L}$ and $\boldsymbol{S}$ around $\boldsymbol{B}$ in the Paschen-Back effect.

We have noted before that optical transitions do not affect the spin, and so we must have $\Delta M_S = 0$. The frequency shift is thus given by:

$$\Delta(h\nu) = \mu_{\rm B} B_z \, \Delta M_L \ , \tag{8.22}$$

where $\Delta M_L = 0$ or ± 1. In other words, we revert to the normal Zeeman effect.

Putting it all together

The change of the spectra as we increase B from zero is illustrated for the p $\rightarrow$ s transitions of an alkali atom in Fig. 8.6. At $B = 0$ the lines are split by the spin-orbit interaction. At weak fields we observe the anomalous Zeeman effect, while at strong fields we change to the Paschen-back effect.

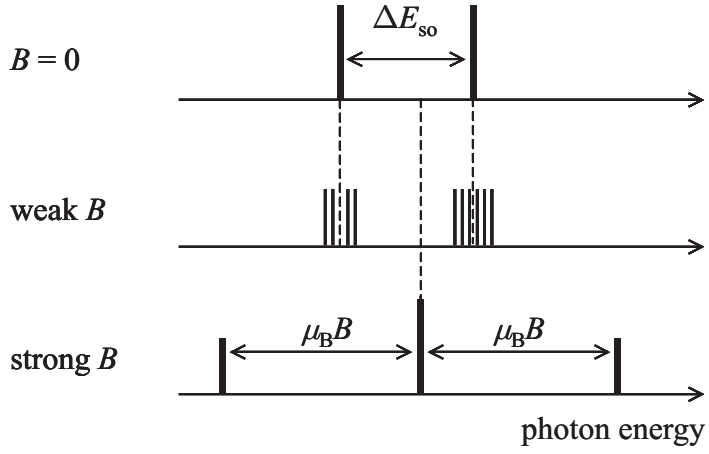

Figure 8.6: Schematic progression of the optical spectra for the p $\rightarrow$ s transitions of an alkali atom with increasing field.

8.1.4 Magnetic field effects for hyperfine levels

Everything we have said so far has ignored the *hyperfine* structure of the atom. The whole process can be repeated to calculate the Zeeman and Paschen-Back energy shifts for the hyperfine levels. In this case, the energy splittings at $B = 0$ are much smaller, due to the much smaller gyromagnetic ratio of the nucleus compared to the electron. (See Section 7.7.2.) This implies that the change from the weak-field to the strong-field limit occurs at much smaller field strengths than for the states split by fine-structure interactions. We shall not consider the hyperfine states further in this course.

8.2 The concept of 'good' quantum numbers

It is customary to refer to quantum numbers that relate to constants of the motion as 'good' quantum numbers. In this discussion of the effects of magnetic fields, we have used six different quantum numbers to describe the angular momentum state of the atom: J, M_J, L, M_L, S, M_S. However, we cannot know

all of these at the same time. In fact, we can only know four: (L, S, J, M_J) in the weak-field limit, or (L, S, M_L, M_S) in the strong-field limit. In the weak-field limit, L_z and S_z are not constant which implies that J and M_J are 'good' quantum numbers but M_L and M_S are not. Similarly, in the strong-field limit, the coupling between $\boldsymbol{L}$ and $\boldsymbol{S}$ is broken and so $\boldsymbol{J}$ and J_z are not constants of the motion: M_L and M_S are good quantum numbers, but J and M_J are not.

A similar type of argument applies to the two angular momentum coupling schemes discussed in Section 5.6, namely LS-coupling and jj-coupling. As an example, consider the total angular momentum state of a two electron atom. In the LS-coupling scheme, we specify (L, S, J, M_J), whereas in the jj-coupling scheme we have (j_1, j_2, J, M_J). In both cases, we have four 'good' quantum numbers, which tell us the precisely measurable quantities. The other quantum numbers are unknown because the physical quantities they represent are not constant. In LS-coupling we cannot know the j values of the individual electrons because the residual electrostatic potential overpowers the spin-orbit effect, whereas in the jj-coupling scheme we cannot know L and S. Note, however, that J and M_J are good quantum numbers in both coupling limits. This means that we can always describe the Zeeman energy of the atom by eqn 8.17, although in the case of jj-coupling, the formula for the g_J factor given in eqn 8.15 will not be valid because L and S are not good quantum numbers.

8.3 Nuclear magnetic resonance

Everything that has been covered so far in this chapter applies to the electrons in the atom. However, a discussion of the Zeeman effect would not be complete if we did not at least mention the interaction of an external magnetic field with the nucleus. As noted in Section 7.7.2, the nucleus has spin, and this gives it a magnetic dipole moment. In analogy with eqn 8.14, the nuclear dipole moment is written (see eqn 7.51):

$$\boldsymbol{\mu}_{\text{nucleus}} = g_I \frac{\mu_{\text{N}}}{\hbar} \boldsymbol{I} \,, \tag{8.23}$$

where g_I is the nuclear g-factor, and $\mu_{\text{N}} = e\hbar/2m_{\text{p}}$ is the nuclear magneton. $\boldsymbol{I}$ is the nuclear spin angular momentum, which is assumed to be quantised in the usual way, so that $|\boldsymbol{I}| = \sqrt{I(I+1)}\hbar$ and $I_z = M_I\hbar$, with M_I running in integer steps from $-I$ to $+I$. Note that the omission of the minus sign in eqn 8.23 is deliberate, as nuclei are positively charged.

If an external magnetic field is applied along the z direction, the energy of the nucleus will shift by:

$$\Delta E = -\boldsymbol{\mu}_{\text{nucleus}} \cdot \boldsymbol{B} = -\mu_z^{\text{nucleus}} B_z \,. \tag{8.24}$$

On substituting from eqn 8.23, the Zeeman energy becomes:

$$\Delta E = -g_I \frac{\mu_{\text{N}}}{\hbar} I_z B_z = -g_I \mu_{\text{N}} B_z M_I \,. \tag{8.25}$$

In magnetic resonance experiments, a radio-frequency (RF) electromagnetic field is applied to induce magnetic-dipole transitions between the Zeeman-split levels. The angular momentum of the nucleus changes by one unit when the photon is absorbed, so that the selection rule is $\Delta M_I = \pm 1$.[4] The energy of the photon required to induce this transition is thus given by:

$$h\nu = g_I \mu_{\text{N}} B_z \,. \tag{8.26}$$

The resonance is detected either by scanning ν at fixed B_z, or by scanning B_z at fixed ν.

In the magnetic resonance systems used in medical imaging, the RF photons are brought to resonance with the hydrogen atoms or ions in the body. The g factor of the proton is 5.586, which implies that $\nu = 42.6\,\text{MHz}$ at a field of 1 T. The non-obvious value of the g value is a consequence of the internal structure of the proton. Magnetic resonance can also be observed from other nuclei in a variety of liquid and solid-state environments, and this gives rise to a host of techniques used especially in chemistry and biology to obtain information about the structure and bonding of molecules.

8.4 Electric fields

In the case of *electric* fields, the weak and strong field limits are not normally distinguished, and all the phenomena are collectively called the **Stark effect**. These effects are named after J. Stark, who was the

[4] In magnetic-dipole transitions, the parity of the initial and final states does not change. (See section 3.5.) The photon interacts with the magnetic dipole of the nucleus, since its electric dipole is zero.

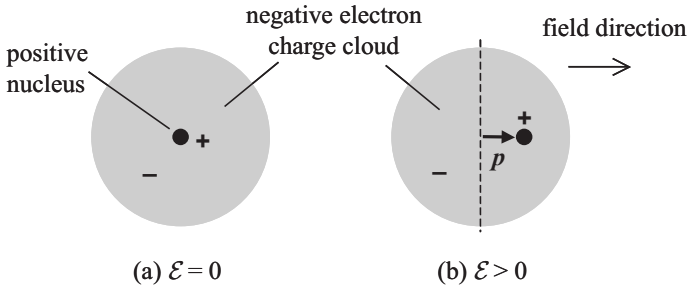

Figure 8.7: Effect of an electric field $\mathcal{E}$ on the electron cloud of an atom. (a) When $\mathcal{E} = 0$, the negatively-charged electron cloud is arranged symmetrically about the nucleus, and there is no electric dipole. (b) When the electric field is applied, the electron cloud is displaced, and a net dipole parallel to the field is induced.

first person to study the effect of electric fields on atomic spectral lines, when he measured the splitting of the hydrogen Balmer lines in an electric field in 1913. In most atoms we observe the **quadratic Stark effect** and we therefore consider this effect first. We then move on to consider the **linear Stark effect**, which is observed for the excited states of hydrogen, and in other atoms at very strong fields. The Stark shift of an atom is harder to observe than the Zeeman shift, which explains why magnetic effects are more widely studied in atomic physics. However, large Stark effects are readily observable in solid state physics, and we therefore conclude by briefly considering the **quantum-confined Stark effect**.

8.4.1 The quadratic Stark effect

Most atoms show a small *red shift* (i.e. a shift to lower energy) which is proportional to the *square* of the electric field. This phenomenon is therefore called the *quadratic* Stark effect. The energy of an atom in an electric field $\boldsymbol{\mathcal{E}}$ is given by

$$E = -\boldsymbol{p} \cdot \boldsymbol{\mathcal{E}}\,, \tag{8.27}$$

where $\boldsymbol{p}$ is the electric dipole of the atom. We can understand the quadratic Stark effect intuitively with reference to Fig. 8.7. The negatively-charged electron clouds of an atom are spherically symmetric about the positively-charged nucleus in the absence of applied fields. A charged sphere acts like a point charge at its centre, and it is thus apparent that atoms do not normally possess a dipole moment, as shown in Fig. 8.7(a). When a field is applied, the electron cloud and the nucleus experience opposite forces, which results in a net displacement of the electron cloud with respect to the nucleus, as shown in Fig. 8.7(b). This creates a dipole $\boldsymbol{p}$ which is parallel to $\mathcal{E}$ and whose magnitude is proportional to $|\boldsymbol{\mathcal{E}}|$. This can be expressed mathematically by writing:

$$\boldsymbol{p} = \alpha \boldsymbol{\mathcal{E}}\,, \tag{8.28}$$

where α is the **polarizability** of the atom. The energy shift of the atom is found by calculating the energy change on increasing the field strength from zero:

$$\Delta E = -\int_0^{\mathcal{E}} \boldsymbol{p}\cdot \mathrm{d}\boldsymbol{\mathcal{E}}' = -\int_0^{\mathcal{E}} \alpha \mathcal{E}' \mathrm{d}\mathcal{E}' = -\frac{1}{2}\alpha \mathcal{E}^2\,, \tag{8.29}$$

which predicts a quadratic red shift, as required. The magnitude of the red shift is generally rather small. This is because the electron clouds are tightly bound to the nucleus, and it therefore requires very strong electric fields to induce a significant dipole.

A more detailed description of the quadratic Stark effect is given in § D.1 of Appendix D.[5] It is shown there that the energy shift of the ith state is given by 2nd order perturbation theory as:

$$\Delta E_i = \sum_{j \neq i} \frac{|\langle \psi_i | H' | \psi_j \rangle|^2}{E_i - E_j}\,, \tag{8.30}$$

where the summation runs over all the other states of the system, and E_i and E_j are the unperturbed energies of the states. Explicit evaluation of the matrix elements for sodium indicates that the Stark shift

[5] This analysis is relegated to the appendix because many of you will not have done perturbation theory yet, as it is normally first encountered in detail in course PHY309, which is taken in the second semester.

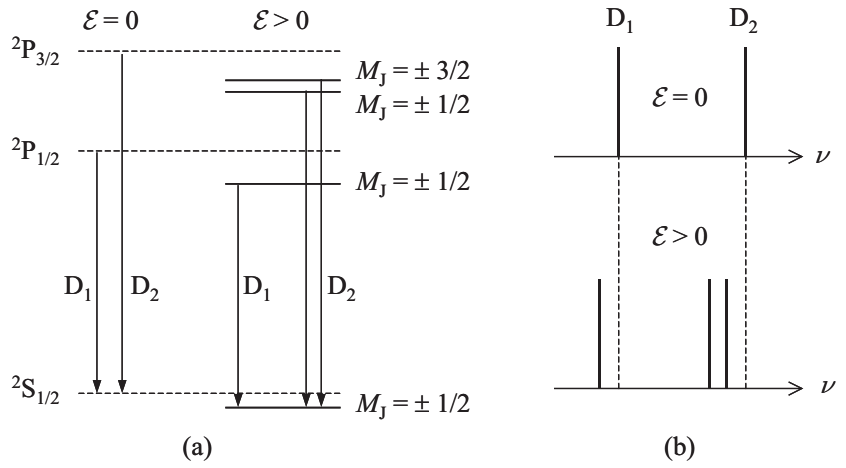

Figure 8.8: (a) Shift of the $^2S_{1/2}$, $^2P_{1/2}$, and $^2P_{3/2}$ terms of an alkali atom in an electric field. Note that the red shifts of the upper levels are larger than that of the lower level. (b) Red shift of the D_1 ($^2P_{1/2} \rightarrow {}^2S_{1/2}$) and the D_2 ($^2P_{3/2} \rightarrow {}^2S_{1/2}$) lines in the field.

at a given field strength depends on M_J^2. This means that electric fields do not completely break the degeneracy of the M_J sub-levels of a particular $|J\rangle$ level, which contrasts with the Zeeman effect, where the energy shift is proportional to M_J, and the degeneracy if fully lifted.

The quadratic Stark shift of the sodium D lines is shown schematically in Fig. 8.8. All states are shifted to lower energy, with those of the same M_J values being shifted equally for a given level, as indicated in Fig. 8.8(a). As discussed in § D.1, the shifts of the upper 3p levels are larger than that of the lower 3s $^2S_{1/2}$ term, and both spectral lines therefore show a net shift to lower energy, as indicated in Fig. 8.8(b). Owing to the degeneracy of the sub-levels with the same $|M_J|$, the D_1 ($^2P_{1/2} \rightarrow {}^2S_{1/2}$) line does not split, while the D_2 ($^2P_{3/2} \rightarrow {}^2S_{1/2}$) line splits into a doublet.

An interesting consequence of the perturbation caused by the electric field is that the unperturbed atomic states get mixed with other states of the opposite parity. For example, the 3s state has even parity at $\mathcal{E} = 0$, but acquires a small admixture of the odd parity 3p state as the field is increased. This means that parity forbidden transitions (eg s→s, p→p, d→s, etc.) become weakly allowed as the field is increased. Since we are dealing with a second-order perturbation, the intensity of these forbidden transitions increases in proportion to $\mathcal{E}^2$.

8.4.2 The linear Stark effect

Stark's original experiment of 1913 was performed on the Balmer lines of hydrogen.[6] In contrast to what has been discussed in the previous subsection, the shift was quite large, and varied linearly with the field. The explanation of the linear Stark effect in hydrogen by degenerate perturbation theory is given in § D.2 of Appendix D. It is shown there that the linear Stark effect is observed when an atom possesses degenerate states of opposite parities. The classic example is the 2s and 2p states of hydrogen, that are degenerate in the absence of spin-orbit effects. Perturbation theory predicts that the $n = 2$ shell splits into a triplet, with energies of $-3ea_0\mathcal{E}$, 0, and $+3ea_0\mathcal{E}$ with respect to the unperturbed level, where a_0 is the Bohr radius of hydrogen. The splitting is linear in the field and has a much larger magnitude than that calculated for the quadratic Stark effect. For example, at $\mathcal{E} = 2.5 \times 10^7\,\mathrm{V/m}$, we find shifts of $\pm 4 \times 10^{-3}\,\mathrm{eV}$ ($32\,\mathrm{cm}^{-1}$), which are more than two orders of magnitude larger than the shifts of the levels in sodium at the same field strength. This, of course, explains why the linear Stark effect in hydrogen was the first electric-field induced phenomenon to be discovered.

The second-order perturbation analysis discussed in § D.1 is expected to break down at large field strengths when the field-induced perturbation becomes comparable to the splittings of the unperturbed levels. The field strength required to reach this limit for the 3s ground-state level of sodium was shown in § D.1 to be extremely large ($\sim 10^{10}\,\mathrm{V/m}$). However, the fields required for the breakdown of the second-order approach for the excited states can be significantly smaller, because some atoms can have different parity excited states which are relatively close to each other. We would then expect the behaviour to

[6] The Balmer series of hydrogen corresponds to those lines that terminate on the $n = 2$ level. These lines occur in the visible spectral region.

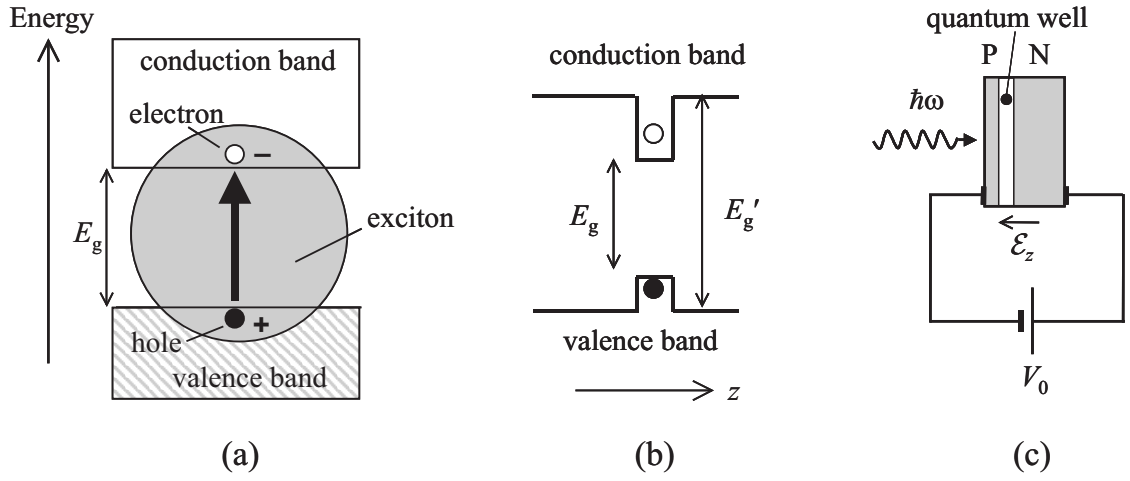

Figure 8.9: The quantum confined Stark effect. (a) Excitons are created by optical transitions from the valence to the conduction band of a semiconductor. (b) A quantum well is formed when a thin layer of a semiconductor with a band gap E_g is sandwiched between layers of another semiconductor with a larger band gap E_g'. (c) Electric fields are applied to an exciton in a quantum well by embedding the quantum well within a P-N junction and applying reverse bias.

change as the field is increased. At low fields we would observe the quadratic Stark effect, but when the field is sufficiently large that the perturbation is comparable to the energy splitting, we would effectively have degenerate levels with different parities, giving rise to a *linear* shift determined by degenerate perturbation theory. This change from the quadratic to linear Stark effect at high fields was first studied for the excited states of helium by Foster in 1927.

8.4.3 The quantum-confined Stark effect

An optical transition between the valence and conduction bands of a semiconductor leaves a positively-charged hole in the valence band, and a negatively-charged electron in the conduction band, as shown in Fig. 8.9(a). The electron and hole can bind together to form a hydrogen-like atom called an **exciton**. The binding energy of the exciton is rather small, due to the high relative dielectric constant ϵ_r of the semiconductor, and also because of the low reduced effective mass of the exciton. Typical values might be $\epsilon_r \sim 10$ and $m \sim 0.1 m_e$, which implies from eqn 2.6 that the 1s binding energy would be $\sim 0.01\,\text{eV}$.[7]

From the discussion given in Section 8.4.1, we would expect the 1s exciton state to show a quadratic Stark shift as an electric field is applied. However, in bulk semiconductors the excitons are very unstable to applied electric fields due to their low binding energy, which implies that the electrostatic force between the electron and hole is relatively small. The electrons and holes are pushed in opposite directions, and the exciton then easily gets ripped apart by the field. This effect is called **field ionization**. It can also be observed in atomic physics, but only at extremely high field strengths.

The situation in a *quantum-confined* structure such as a semiconductor **quantum well** or **quantum dot** is rather different. Consider the case of the quantum well shown in Fig. 8.9(b). The quantum well is formed by sandwiching a thin semiconductor with a band gap of E_g between layers of another semiconductor with a larger band gap E_g'. This then gives rise to spatial discontinuities in the conduction and valence band energies as shown in the figure. The excitons that are formed by optical transitions across the smaller band gap are then trapped in the finite potential well created by the band discontinuities.

A strong electric field can be applied to the quantum well by embedding it within a P-N junction, and then applying reverse bias, as shown in Fig. 8.9(c). P-N junctions conduct when forward bias is applied, but not under reverse bias. In the latter case, the applied voltage is dropped across the narrow junction region, thereby generating an electric field that is controlled by the reverse bias. The excitons that are created by optical transitions are now stable to the field, because the barriers of the quantum well prevent them from being ripped apart. The electrons are pushed to one side of the quantum well, and the holes to the other, which creates a dipole of magnitude $\sim ed$, where d is the width of the quantum well. With

[7] Note that the factor of ϵ_0^2 in the denominator of eqn 2.6 has to be replaced by $(\epsilon_r \epsilon_0)^2$ in a dielectric medium.

$d \sim 10\,\mathrm{nm}$, much larger dipoles can be created than in atomic physics, resulting in correspondingly larger Stark shifts. This effect is called the **quantum-confined Stark effect**, and is widely used for making electro-optical modulators. The quantum-confined Stark effect will be studied in more detail in course PHY475.

Reading

Bransden and Joachain, *Physics of Atoms and Molecules*, chapter 6, §9.8–9, 16.1
Beisser, *Concepts of Modern Physics*, section 6.10.
Demtröder, *Atoms, Molecules and Photons*, sections 5.2, 5.6, 7.2 and 11.9.
Eisberg and Resnick, *Quantum Physics*, section 10.6.
Foot, *Atomic Physics*, sections 1.8 and 5.5
Haken and Wolf, *The physics of atoms and quanta*, chapters 13 and 15, §20.6–7
Hertel and Schulz, *Atoms, Molecules and Optical Physics, 1*, §8.1 – 8.2, 9.5

Chapter 9

Lasers I: Stimulated emission

9.1 Introduction

The word "LASER" is an acronym standing for "Light Amplification by Stimulated Emission of Radiation". The origins of the laser may be traced back to Einstein's seminal paper on stimulated emission published in 1917, but it took until 1960 for the first laser to be invented. It is difficult to identify all of the key milestones in the history of laser physics, but here are a few of the more important ones:

1917 Einstein's treatment of stimulated emission.

1951 Development of the maser by C.H. Townes.[1]

1958 Proposal by C.H. Townes and A.L. Schawlow that the maser concept could be extended to optical frequencies.

1960 T.H. Maiman at Hughes Laboratories reports the first laser: the pulsed ruby laser.

1961 The first continuous wave laser is reported: the helium neon laser.

1962 Invention of the semiconductor laser.

1964 Nicolay Basov, Charlie Townes and Aleksandr Prokhorov are awarded the Nobel prize for "fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser-laser principle."

1981 Art Schalow and Nicolaas Bloembergen are awarded the Nobel Prize for "their contribution to the development of laser spectroscopy."

1997 Steven Chu, Claude Cohen-Tannoudji and William D. Phillips are awarded the Nobel Prize for the "development of methods to cool and trap atoms with laser light."

2005 John Hall and Theodor Hänsch receive the Nobel Prize for "their contributions to the development of laser-based precision spectroscopy, including the optical frequency comb technique".

2010 50th anniversary of the laser.

Many different types of laser have been developed over the years. The "L" in "laser" stands for "Light", but light is understood here in a general sense to mean electromagnetic radiation with a frequency of $\sim 10^{14}$–10^{15} Hz, not specifically visible radiation. This provides the first general classification of laser types:

- infrared, visible or ultraviolet wavelength.

Other general classifications include:

[1] A maser is basically the same as a laser, except that it works at microwave rather than optical frequencies. It took some years to move on from masers to lasers because microwave cavities are designed on the assumption that the cavity dimensions are comparable to the wavelength of the radiation within the cavity, which is typically around 10 cm at microwave frequencies. Such designs cannot be scaled easily to optical wavelengths, where $\lambda \sim 1\,\mu$m, and it required some lateral thinking to design a cavity that would work in the regime where the cavity dimensions are much larger than the wavelength. It is only relatively recently that it has been possible to make "microcavity lasers" and "nanolasers" that have physical dimensions that are comparable to the wavelength of light.

- solid, liquid or gas gain medium;
- continuous wave (CW) or pulsed operation;
- fixed wavelength or tuneable wavelength.

The gain medium (i.e. amplifying medium) of the laser determines the possible wavelengths that the laser can emit, but the characteristic properties of laser light are also strongly affected by the design of the cavity, which is the other essential part of a laser. Such properties include:

Monochromaticity Discharge lamps emit light of many different colours simultaneously, according to the emission probabilities of the transitions in the atoms. Lasers, by contrast, emit light from just a single atomic transition, and are therefore highly monochromatic.[2] The transition that is "selected" by the laser is determined by the amount of amplification that is available at that wavelength and the reflectivity of the mirrors that comprise the cavity.

Directionality Discharge lamps emit in all directions, but lasers emit a well-defined beam in a specific direction. The direction of the beam is governed by the orientation of the mirrors in the cavity.

Brightness The brightness of lasers arises from two factors. First, the radiation is emitted in a beam, which means that the intensity (i.e. power per unit area) can be very high, even though the total amount of power is relatively low. Second, all of the energy is concentrated within the narrow spectrum of a single atomic transition. This means that the **spectral brightness** (i.e. the intensity in the beam divided by the spectral width of the emission line) is even higher in comparison with a white-light source such as an incandescent light bulb. For example, the spectral brightness of a 1 mW laser beam could easily be millions of time greater than that of a 100 W light bulb.

Coherence Lasers have a high degree of both **spatial** and **temporal** coherence. The coherence of laser light will be considered in more detail in Section 10.3.

These four properties are common to all lasers. In addition, some lasers emit radiation in very short pulses, which can be used for studying fast processes in physics, chemistry, and biology, or for transmitting optical data at a very high rate down optical fibres. The principles that govern whether a laser can produce very short pulses are considered in Section. 10.2.2.

9.2 Principles of laser oscillation

As mentioned above, the word "LASER" is an acronym that stands for "light amplification by stimulated emission of radiation". However, there is more to a laser than just light amplification. A laser is actually an oscillator rather than just an amplifier.[3] The difference is that an oscillator has positive feedback in addition to amplification. The key ingredients of a laser may thus be summarized as:

$$\boxed{\text{LASER} = \text{light amplification} + \text{positive optical feedback}}\,.$$

Light amplification is achieved by **stimulated emission**. Ordinary optical materials do not amplify light. Instead, they tend to absorb or scatter the light, so that the light intensity out of the medium is less than the intensity that went in. To get amplification you have to drive the material into a non-equilibrium state by pumping energy into it. The amplification of the medium is determined by the **gain coefficient** γ, which is defined by the following equation:

$$I(x + \mathrm{d}x) = I(x) + \gamma I(x)\mathrm{d}x \equiv I(x) + \mathrm{d}I\,, \tag{9.1}$$

where $I(x)$ represents the intensity at a point x within the gain medium. The differential equation can be solved as follows:

$$\begin{aligned} \mathrm{d}I &= \gamma I \mathrm{d}x \\ \therefore \frac{\mathrm{d}I}{\mathrm{d}x} &= \gamma I \\ \therefore I(x) &= I(0)\mathrm{e}^{\gamma x}\,. \end{aligned} \tag{9.2}$$

[2] "Monochromatic" means "single coloured".

[3] A more accurate acronym for a laser might therefore be "LOSER", but it is easy to understand why this one never caught on.

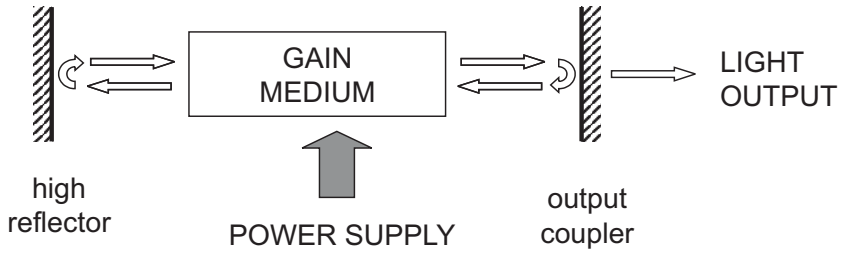

Figure 9.1: Schematic diagram of a laser

Thus the intensity grows exponentially within the gain medium.

Positive optical feedback is achieved by inserting the amplifying medium inside a resonant **cavity**. The effects of the cavity on the properties of laser light will be considered in detail in Sections 10.1–10.3. At this stage, we confine ourselves to considering the parameters of the cavity that affect the condition for laser oscillation.

Figure 9.1 shows a schematic diagram of a laser. Light in the cavity passes through the gain medium and is amplified. It then bounces off the end mirrors and passes through the gain medium again, getting amplified further. This process repeats itself until a stable equilibrium condition is achieved when the total round trip gain balances all the losses in the cavity. Under these conditions the laser will oscillate. The condition for oscillation is thus:

$$\boxed{\text{round-trip gain} = \text{round-trip loss}} \quad .$$

The losses in the cavity fall into two categories: useful, and useless. The useful loss comes from the output coupling. One of the mirrors (called the **output coupler**) has reflectivity less than unity, and allows some of the light oscillating around the cavity to be transmitted as the output of the laser. The value of the transmission is chosen to maximize the output power. If the transmission is too low, very little of the light inside the cavity can escape, and thus we get very little output power. On the other hand, if the transmission is too high, there may not be enough gain to sustain oscillation, and there would be no output power. The optimum value is somewhere between these two extremes. Useless losses arise from absorption in the optical components (including the laser medium), scattering, and the imperfect reflectivity of the other mirror (the **high reflector**). By taking into account the fact that the light passes twice through the gain medium during a round trip, the condition for oscillation in a laser can be written:

$$\mathrm{e}^{2\gamma l}\, R_{\mathrm{OC}}\, R_{\mathrm{HR}}\, \mathbb{L} = 1\,, \tag{9.3}$$

where l is the length of the gain medium, R_{OC} is the reflectivity of the output coupler, R_{HR} is the reflectivity of the high reflector, and $\mathbb{L}$ is the round-trip loss factor due to absorption and scattering, such that $\mathbb{L} = 1$ corresponds to the situation with no losses. If the total round-trip losses are small ($\lesssim 10\%$), then the gain required to sustain lasing will also be small, and eqn 9.3 simplifies to:

$$2\gamma l = (1 - R_{\mathrm{OC}}) + (1 - R_{\mathrm{HR}}) + \text{scattering losses} + \text{absorption losses}\,. \tag{9.4}$$

This shows more clearly how the gain in the laser medium must exactly balance the losses in the cavity.

In general we expect the gain to increase as we pump more energy into the laser medium. At low pump powers, the gain will be small, and there will be insufficient gain to reach the oscillation condition. The laser will not start to oscillate until there is enough gain to overcome all of the losses. This implies that the laser will have a **threshold** in terms of the pump power. (See Section 9.6.)

9.3 Stimulated emission

In Chapter 3, we considered the spontaneous tendency for atoms in excited states to emit radiation. We now consider the optical transitions that occur when the atom is subjected to electromagnetic radiation with its frequency resonant with the energy difference of the two levels. We follow the treatment of Einstein (1917).

In addition to transitions from the upper to the lower level due to **spontaneous emission**, there will also be:

- **absorption** of photons causing transitions from level 1 up to level 2;

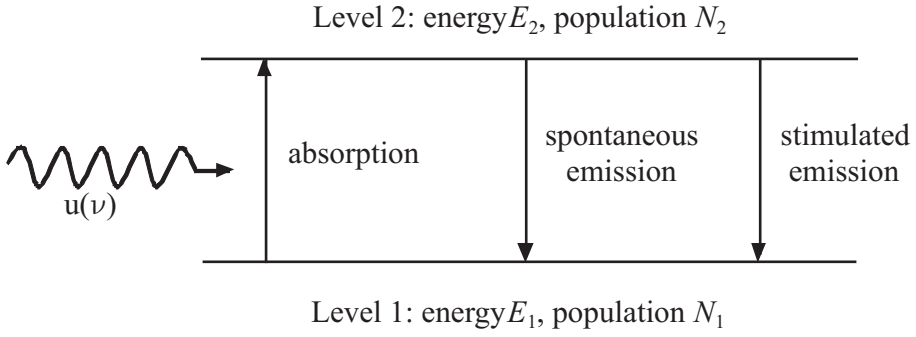

Figure 9.2: Absorption, spontaneous emission, and stimulated emission transitions

- **stimulated emission** in which atoms in level 2 drop to level 1 induced by the incident radiation.

The process of stimulated emission is a coherent quantum-mechanical effect. The photons emitted by stimulated emission are in phase with the photons that induce the transition. This is the fundamental basis of laser operation, as the name suggests: Light Amplified by *Stimulated Emission* of Radiation.

Consider an atom irradiated by white light, with N_2 atoms in level 2 and N_1 atoms in level 1. The part of spectrum at frequency ν, where $h\nu = (E_2 - E_1)$, can induce absorption and stimulated emission transitions. We write the spectral energy density of the light at frequency ν as $u(\nu)$. The transitions that can occur are shown in Fig. 9.2. In order to treat this situation, Einstein introduced his A and B coefficients. We have already seen in Section 3.6 that the A coefficient determines the rates of spontaneous transitions. The introduction of the B coefficient extends the treatment to include absorption and stimulated emission. The transition rates for three processes are:

- Spontaneous emission ($2 \to 1$):
$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\frac{\mathrm{d}N_1}{\mathrm{d}t} = -A_{21}N_2\,. \tag{9.5}$$
- Stimulated emission ($2 \to 1$):
$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\frac{\mathrm{d}N_1}{\mathrm{d}t} = -B_{21}N_2u(\nu)\,. \tag{9.6}$$
- Absorption ($1 \to 2$):
$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = -\frac{\mathrm{d}N_2}{\mathrm{d}t} = -B_{12}N_1u(\nu)\,. \tag{9.7}$$

These are effectively the definitions of the Einstein A and B coefficients. At this stage we might be inclined to think that the three coefficients are independent parameters for a particular transition. This is not in fact the case. If you know one, you can work out the other two. To see this, we follow Einstein's analysis.

We imagine a gas of atoms inside a box at temperature T with black walls. If we leave the atoms for long enough, they will come to equilibrium with the black-body radiation that fills the cavity. In these steady-state conditions, the rate of upward transitions must exactly balance the rate of downward transitions:
$$B_{12}N_1u(\nu) = A_{21}N_2 + B_{21}N_2u(\nu)\,, \tag{9.8}$$
which implies that:
$$\frac{N_2}{N_1} = \frac{B_{12}u(\nu)}{A_{21} + B_{21}u(\nu)}\,. \tag{9.9}$$
Furthermore, since the gas is in thermal equilibrium at temperature T, the ratio of N_2 to N_1 must satisfy Boltzmann's law:
$$\frac{N_2}{N_1} = \frac{g_2}{g_1}\exp\left(-\frac{h\nu}{k_\mathrm{B}T}\right)\,, \tag{9.10}$$
where g_2 and g_1 are the degeneracies of levels 2 and 1 respectively, and $h\nu = (E_2 - E_1)$. Equations 9.9 and 9.10 together imply that:
$$\frac{B_{12}u(\nu)}{A_{21} + B_{21}u(\nu)} = \frac{g_2}{g_1}\exp\left(-\frac{h\nu}{k_\mathrm{B}T}\right)\,. \tag{9.11}$$

On solving this for $u(\nu)$, we find:

$$u(\nu) = \frac{g_2 A_{21}}{g_1 B_{12} \exp(h\nu/k_{\rm B}T) - g_2 B_{21}} \,. \tag{9.12}$$

However, we know that the cavity is filled with black-body radiation, which has a spectral energy density given by the Planck formula:

$$u(\nu) = \frac{8\pi h\nu^3}{(c/n)^3} \frac{1}{\exp(h\nu/k_{\rm B}T) - 1} \,, \tag{9.13}$$

where c/n is the speed of light in a medium with refractive index n. The only way to make eqns 9.12 and 9.13 to be consistent with each other at all temperatures and frequencies is if:

$$\begin{aligned} g_1 B_{12} &= g_2 B_{21} \,, \\ A_{21} &= \frac{8\pi n^3 h\nu^3}{c^3} B_{21} \,. \end{aligned} \tag{9.14}$$

A moment's thought will convince us that it is not possible to get consistency between the equations without the stimulated emission term. This is what led Einstein to introduce the concept.

The relationships between the Einstein coefficients given in eqn 9.14 have been derived for the case of an atom in thermal equilibrium with black-body radiation at temperature T. However, once we have derived the inter-relationships, they will apply in all other cases as well. This is very useful, since we only need to know one of the coefficients to work out the other two. For example, we can measure the radiative lifetime to determine A_{21} from (cf. eqn 3.25),

$$A_{21} = \frac{1}{\tau} \,, \tag{9.15}$$

and then work out the B coefficients from eqn 9.14.

Equation 9.14 shows that the probabilities for absorption and emission are the same apart from the degeneracy factors, and that the ratio of the probability for spontaneous emission to stimulated emission increases in proportion to ν^3. In a laser we want to encourage stimulated emission and suppress spontaneous emission. Hence it gets progressively more difficult to make lasers work as the frequency increases, all other things being equal.

9.4 Population inversion

We have seen above that stimulated emission is the basis of laser operation. We now wish to study how we can use stimulated emission to make a light amplifier. In a gas of atoms in thermal equilibrium, the population of the lower level will always be greater than the population of the upper level. (See eqn 9.10). Therefore, if a light beam is incident on the medium, there will always be more upward transitions due to absorption than downward transitions due to stimulated emission. Hence there will be net absorption, and the intensity of the beam will diminish on progressing through the medium.

To amplify the beam, we require that the rate of stimulated emission transitions exceeds the rate of absorption.[4] We see from eqns 9.6 and 9.7 that this condition is satisfied when:

$$B_{21} N_2 u(\nu) > B_{12} N_1 u(\nu) \,. \tag{9.16}$$

On substituting from eqn 9.14, this leads to the conclusion:

$$N_2 > \frac{g_2}{g_1} N_1 \,, \tag{9.17}$$

which simplifies to:

$$N_2 > N_1 \,, \tag{9.18}$$

for non-degenerate levels. This is a highly non-equilibrium situation, and is called **population inversion**. On comparing eqn 9.17 to eqn 9.10, we see that population inversion corresponds to *negative* temperatures. This is not as ridiculous as it sounds, because the atoms are not in thermal equilibrium.

Once we have population inversion, we have a mechanism for generating gain in the laser medium. The art of making a laser is to work out how to get population inversion for the relevant transition. We will discuss how this is done for specific types of laser in Sections 10.5–10.6.

[4] We can usually ignore spontaneous emission in an operating laser because we are considering the case in which the light intensity is very high, so that the stimulated emission rate far exceeds the spontaneous emission rate.

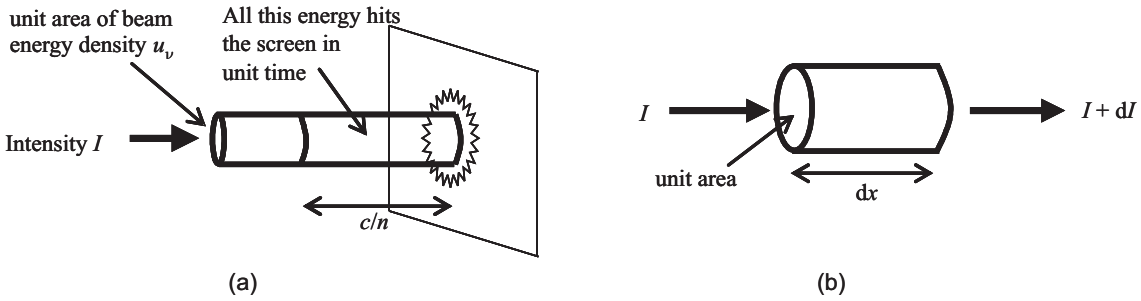

Figure 9.3: (a) Relationship between the intensity I and energy density u_ν of a light beam. (b) Incremental intensity increase in a gain medium.

9.5 Gain coefficient

Having realized that population inversion gives rise to amplification of light, we now want to work out a relationship between the gain coefficient and the population inversion density. Before we can proceed, we must first refine our analysis of the absorption and stimulated emission rates. Einstein's analysis considered the interaction of an ideal atom with a featureless white-light spectrum. In practice, we are more interested in the interaction of real atoms with sharp emission lines with an even narrower band of light that will eventually become the laser mode.

The energy density $u(\nu)$ that appears in eqns 9.6–9.13 is the *spectral* energy density (units: $\mathrm{J\,m^{-3}\,Hz^{-1}} \equiv \mathrm{J\,s\,m^{-3}}$). We now consider the interaction between an atom with a normalized line shape function $g(\nu)$ as defined in Section 3.7 (units: $\mathrm{Hz^{-1}}$) and a beam of light whose emission spectrum is much narrower than the spectral line width of the atomic transition. This situation is considered in Appendix E, where it is shown that the rates of absorption and stimulated emission can be written respectively as: (cf. eqns E.4 and E.5 with the subscript on the laser frequency omitted):

$$\begin{aligned} W_{12} &= B_{12} N_1 u_\nu g(\nu)\,, \\ W_{21} &= B_{21} N_2 u_\nu g(\nu)\,. \end{aligned} \tag{9.19}$$

The light source is considered to have a delta function spectrum at frequency ν with *total* energy density u_ν per unit volume (units $\mathrm{J\,m^{-3}}$). u_ν is related to the intensity I of the optical beam by (see Fig. 9.3(a)):

$$I = u_\nu \frac{c}{n}\,, \tag{9.20}$$

where n is the refractive index of the medium. This means that the net stimulated rate downwards from level 2 to level 1 is given by:

$$W_{21}^{\mathrm{net}} \equiv W_{21} - W_{12} = (N_2 - N_1) B_{21} g(\nu) \frac{n}{c} I\,, \tag{9.21}$$

where we have assumed that the levels are non-degenerate so that $B_{12} = B_{21}$.[5] (See eqn 9.14.)

For each net transition a photon of energy $h\nu$ is added to the beam. The energy added to a unit volume of beam per unit time is thus $W_{21}^{\mathrm{net}} h\nu$. Consider a small increment of the light beam inside the gain medium with length $\mathrm{d}x$, as shown in Fig. 9.3(b). The energy added to this increment of beam per unit time is $W_{21}^{\mathrm{net}} h\nu \times \mathbb{A}\,\mathrm{d}x$, where $\mathbb{A}$ is the beam area. On remembering that the intensity equals the energy per unit time per unit area, we can write:

$$\begin{aligned} \mathrm{d}I &= W_{21}^{\mathrm{net}} h\nu\,\mathrm{d}x\,, \\ &= (N_2 - N_1) B_{21} g(\nu) \frac{n}{c} I h\nu\,\mathrm{d}x\,. \end{aligned} \tag{9.22}$$

On comparing this to eqn 9.2, we see that the gain coefficient γ is given by:

$$\gamma(\nu) = (N_2 - N_1) B_{21} g(\nu) \frac{n}{c} h\nu\,. \tag{9.23}$$

[5] For non-degenerate levels, eqn 9.21 gets modified to:

$$W_{21}^{\mathrm{net}} = \left(N_2 - \frac{g_2}{g_1} N_1 \right) B_{21} g(\nu) \frac{n}{c} I \equiv \Delta N B_{21} g(\nu) \frac{n}{c} I\,,$$

where $\Delta N = (N_2 - (g_2/g_1)N_1)$ is the population inversion density for a system with non-dengerate levels. See eqn 9.17.

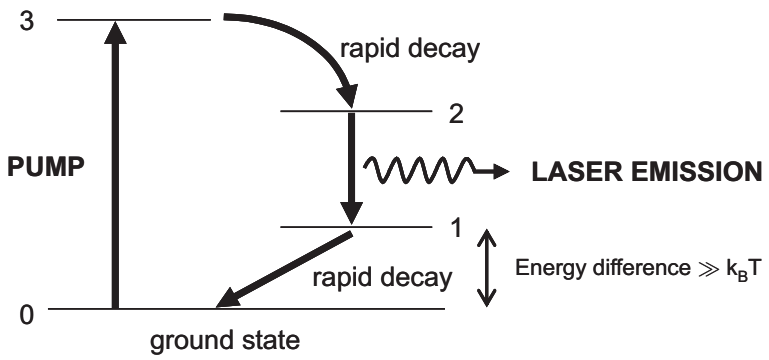

Figure 9.4: Level scheme for a four-level laser.

This result shows that the gain is directly proportional to the population inversion density, and also follows the spectrum of the emission line. By using eqn 9.14 to express B_{21} in terms of A_{21}, we can re-write the gain coefficient in terms of the natural radiative lifetime τ using eqn 9.15 to obtain:

$$\gamma(\nu) = (N_2 - N_1)\frac{\lambda^2}{8\pi n^2 \tau} g(\nu)\,, \tag{9.24}$$

where λ is the vacuum wavelength of the emission line. This is the required result. Equation 9.24 tells us how to relate the gain in the medium to the population inversion density using experimentally measurable parameters: λ, τ, n and $g(\nu)$.

9.6 Laser threshold

Equation 9.24 shows us that the gain in a laser medium is directly proportional to the population inversion density. Laser operation will occur when there is enough gain to overcome the losses in the cavity. This implies that a minimum amount of population inversion must be obtained before the laser will oscillate. Population inversion is achieved by "pumping" atoms into the upper laser level. This pumping can be done by a variety of techniques, which will be described in more detail when we consider individual laser systems in Chapter 10. At this stage we just consider the general principles.

Lasers are classified as being either **three-level** of **four-level** systems. We will consider four-level lasers first, as these are the most common. Examples are Helium Neon or Nd:YAG. The four levels are: the ground state (0), the two lasing levels (1 & 2), and a fourth level (3) which is used as part of the pumping mechanism. The level scheme for an ideal four-level laser is shown in Fig. 9.4. The feature that makes it a four-level as opposed to a three-level laser is that the lower laser level is at an energy more than $k_{\mathrm{B}}T$ above the ground state. This means that the thermal population of level 1 is negligible, and so level 1 is empty before we turn on the pumping mechanism.

We assume that the atoms are inside a cavity and are being pumped into the upper laser level (level 2) at a constant rate of $\mathbb{R}_2$. This is usually done by exciting atoms to level 3 with a bright flash lamp or by an electrical discharge, and then by a rapid decay to level 2. We can write down the following rate equations for the populations of levels 1 and 2:

$$\begin{aligned} \frac{\mathrm{d}N_2}{\mathrm{d}t} &= -\frac{N_2}{\tau_2} - W_{12}^{\mathrm{net}} + \mathbb{R}_2\,, \\ \frac{\mathrm{d}N_1}{\mathrm{d}t} &= +\frac{N_2}{\tau_2} + W_{12}^{\mathrm{net}} - \frac{N_1}{\tau_1}\,. \end{aligned} \tag{9.25}$$

The various terms allow for:

- spontaneous emission from level 2 to level 1 ($\pm N_2/\tau_2$),
- stimulated transitions from level 2 to level 1 ($\pm W_{21}^{\mathrm{net}}$),
- pumping into level 2 ($\mathbb{R}_2$),
- decay from level 1 to the ground state by radiative transitions and/or collisions (N_1/τ_1).

Note that $W_{21}^{\rm net}$ is the net stimulated transition rate from level 2 to level 1, as given in eqn 9.21. This is equal to the rate of stimulated emission transitions downwards minus the rate of stimulated absorption transitions upwards.

There are two important assumptions implicitly contained in eqn 9.25:

1. There is no pumping into level 1.

2. The only decay route from level 2 is by radiative transitions to level 1 (i.e. there are no non-radiative transitions between level 2 and level 1, and transitions to other levels are not possible).

It may not always be possible to satisfy these assumptions, but it helps if we can. That is why we described the above scenario as an "ideal" four-level laser.

We can re-write eqn 9.21 in the following form:

$$W_{21}^{\rm net} = B_{21} g(\nu) \frac{n}{c} I (N_2 - N_1) \equiv \mathbb{W}(N_2 - N_1)\,, \tag{9.26}$$

where $\mathbb{W} = B_{21} g(\nu) n I/c$, and I is the intensity inside the laser cavity. In steady-state conditions, the time derivatives in eqn 9.25 must be zero. We can thus solve eqn 9.25 for N_1 and N_2 using eqn 9.26 to obtain:

$$\begin{aligned} N_1 &= \mathbb{R}_2 \tau_1\,, \\ N_2 &= \frac{\mathbb{W} N_1 + \mathbb{R}_2}{\mathbb{W} + 1/\tau_2}\,. \end{aligned} \tag{9.27}$$

Therefore the population inversion is given by

$$\Delta N \equiv N_2 - N_1 = \frac{\mathbb{R}_2}{\mathbb{W} + 1/\tau_2} \left(1 - \frac{\tau_1}{\tau_2} \right)\,. \tag{9.28}$$

This shows that the population inversion is directly proportional to the pumping rate into the upper level. Note, however, that it is not possible to achieve population inversion (i.e. $\Delta N > 0$) unless $\tau_2 > \tau_1$. This makes sense if you think about it. Unless the lower laser level empties quickly, atoms will pile up in the lower laser level and this will destroy the population inversion.

Equation 9.28 can be re-written as :

$$\Delta N = \frac{\mathbb{R}}{\mathbb{W} + 1/\tau_2}\,, \tag{9.29}$$

where $\mathbb{R} = \mathbb{R}_2(1 - \tau_1/\tau_2)$. This is the net pumping rate after allowing for the unavoidable accumulation of atoms in the lower level because τ_1 is non-zero. If the laser is below the threshold for lasing, there will be very few photons in the cavity. Therefore, $\mathbb{W}$ will be very small because I is very small: see eqn 9.26 above. The population inversion is simply $\mathbb{R}\tau_2$, and thus increases linearly with the pumping rate. Equation 9.24 implies that the gain coefficient similarly increases linearly with the pumping rate below threshold.

Eventually we will have enough gain to balance the round trip losses. This determines the threshold gain coefficient $\gamma^{\rm th}$ for laser oscillation, as set out in eqn 9.3 or 9.4. From eqn 9.24 we have:

$$\Delta N^{\rm th} = \frac{8\pi n^2 \tau_2}{\lambda^2 g(\nu)} \gamma^{\rm th}\,. \tag{9.30}$$

By combining eqns 9.28 and 9.29 with $\mathbb{W} = 0$ we can work out the pumping rate required to instigate lasing. This is the **threshold** pumping rate. It is given by $\mathbb{R}^{\rm th} = \Delta N^{\rm th}/\tau_2$. All lasers have a threshold. Unless you pump them hard enough, they will not work.

What happens if we increase the pumping rate beyond the threshold value? In steady-state conditions, the gain cannot increase any more, which implies that the population inversion is clamped at the value given by eqn 9.30 even when $\mathbb{R}$ exceeds $\mathbb{R}^{\rm th}$. This is shown in Fig. 9.5(a). What about the power output? We set $\mathbb{W}$ to zero in eqn 9.29 because there was very little light in the cavity below threshold. This is no longer true once the laser starts oscillating. If we are above threshold, ΔN is clamped at the value set by eqn 9.30, and so eqn 9.29 tells us that:

$$\mathbb{W} = \frac{\mathbb{R}}{\Delta N^{\rm th}} - \frac{1}{\tau_2} = \frac{1}{\tau_2} \left(\frac{\mathbb{R}}{\mathbb{R}^{\rm th}} - 1 \right) \quad \ldots \quad \text{for } \mathbb{R} > \mathbb{R}^{\rm th}\,. \tag{9.31}$$

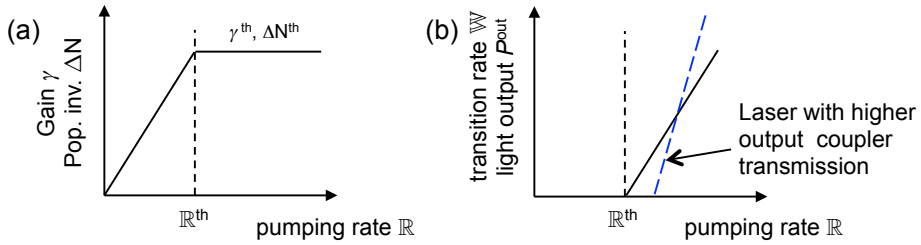

Figure 9.5: (a) Variation of the gain and population inversion in a laser with the pumping rate. (b) Comparison of the threshold and light outputs for two different values of the transmission of the output coupler. Note that these curves only apply to four-level laser systems.

Now $\mathbb{W}$ is proportional to the intensity I inside the cavity (see eqn 9.26), which in turn is proportional to the output power $P^{\rm out}$ emitted by the laser. Thus $P^{\rm out}$ is proportional to $\mathbb{W}$, and we may write:

$$P^{\rm out} \propto \left(\frac{\mathbb{R}}{\mathbb{R}^{\rm th}} - 1 \right) \quad \ldots \quad \text{for } \mathbb{R} > \mathbb{R}^{\rm th} \,. \tag{9.32}$$

This shows that the output power increases linearly with the pumping rate once the threshold has been achieved, as shown in Fig. 9.5(b).

The choice of the reflectivity of the output coupler affects the threshold because it determines the oscillation conditions: see eqn 9.3 or 9.4. If the output coupler transmission $(1 - R_{\rm OC})$ is small, the laser will have a low threshold, but the output coupling efficiency will be low. By increasing the transmission, the threshold increases, but the power is coupled out more efficiently. This point is illustrated in Fig. 9.5(b). The final choice for $R_{\rm OC}$ depends on how much pump energy is available, which will govern the optimal choice to get the maximum output power.

9.7 Pulsed Lasers

So far we have only considered continuous wave (CW) lasers, but many lasers in fact operate in a pulsed mode. Powerful pulsed flash lamps can give rise to very large pumping rates, with correspondingly large output pulse energies, especially when using a trick called "Q-switching". In this technique, the losses in the cavity are kept artificially high by some external method.[6] This prevents lasing and allows the build up of very large population inversion densities, with correspondingly large gain coefficients. If the losses are suddenly reduced, a very powerful pulse will build up because of the very high gain in the cavity. Q-switching is widely used in solid-state lasers because they tend to have long upper state lifetimes, which allows the storage of a large amount of energy in the crystal, but it is seldom used in gas lasers because the lifetimes are shorter which makes it difficult to store much energy in the gain medium.

9.8 Three-level lasers

Some lasers are classified as being **three-level** systems. The standard example is ruby, which was the first laser ever produced. The key difference between a three-level laser and a four-level laser is that the lower laser level is the ground state, as shown in Fig. 9.6(a). On comparing Figs. 9.4 and 9.6, it is apparent that the lower laser level of the four-level system has merged with the ground state in the three-level system. This makes it much more difficult to obtain population inversion in three-level lasers because the lower laser level initially has a very large population.

Consider a system with N_0 atoms. With the pump turned off, all of the atoms will be in the lower laser level, so that $N_1 = N_0$, $N_2 = 0$, and $\Delta N = N_2 - N_1 = -N_0$. By turning on the pump, we excite $\mathrm{d}N$ atoms to level 3, which then decay to level 2. The population N_2 of level 2 is thus $\mathrm{d}N$, while the population N_1 of the lower laser level is $(N_0 - \mathrm{d}N)$. For population inversion we require $N_2 > N_1$ (i.e. ΔN positive), and hence $\mathrm{d}N > (N_0 - \mathrm{d}N)$, which implies $\mathrm{d}N > N_0/2$. Therefore, in order to obtain population inversion we have to pump more than half of the atoms out of the ground state into the upper laser level. This obviously requires a very large amount of energy, which contrasts with four-level lasers,

[6] One way to switch the cavity losses from high to low on fast timescales is to use an electro-optical modulator. This effectively behaves like a fast intra-cavity shutter.

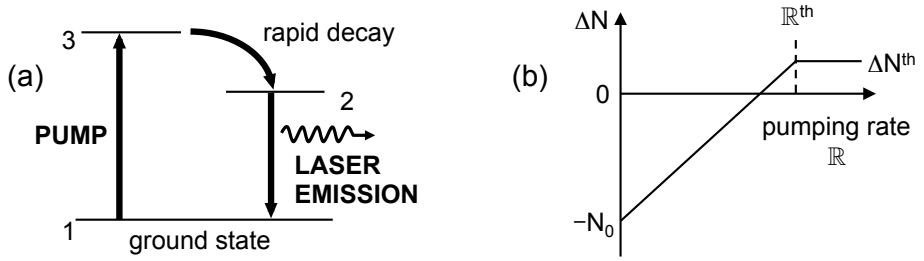

Figure 9.6: (a) Level scheme for a three-level laser, for example: ruby. (b) Variation of the population inversion density ΔN with pumping rate $\mathbb{R}$ in a three-level laser.

where the lower laser level is empty before the pumping process starts, and much less energy is required to reach threshold.

The variation of the inversion density with pumping rate $\mathbb{R}$ for a three-level laser is shown schematically in Fig. 9.6(b). As explained above, the inversion density is equal to $-N_0$ at $\mathbb{R} = 0$, and only becomes positive when more than half of the atoms have been pumped to the upper level. Once ΔN is positive, amplification occurs, and the lasing threshold will be reached when the inversion density is sufficiently large to provide enough gain to overcome the cavity losses. As with the four-level laser, the gain (and hence the inversion density) above threshold are fixed at the level set by the oscillation condition in eqn 9.3, which is first reached at the threshold pumping rate $\mathbb{R}^{\rm th}$.

Despite the fact that the threshold for population inversion is very high in a three-level system, they can be quite efficient once this threshold is overcome. Ruby lasers pumped by bright flash lamps actually give very high output pulse energies. However, they only work in pulsed mode. Continuous lasers tend to be made by using four-level systems.

Further Reading

Bransden and Joachain, *Physics of Atoms and Molecules*, §4.4, 15.1
Demtröder, *Atoms, Molecules and Photons*: §7.1, 8.1
Hecht, *Optics*: §7.4.3, 12.1, 13.1
Hertel and Schulz, *Atoms, Molecules and Optical Physics, 1*, §4.2.3
Hooker and Webb, *Laser Physics*: chapters 1, 2, 4, §6.5–7
Silfvast, *Laser Fundamentals*: Chapter 1, chapter 6–9
Smith and King, *Optics and Photonics*: chapters 15, 17
Svelto, *Principles of Lasers*: Chapter 1, §2.1–4, §7.1–3
Wilson and Hawkes, *Optoelectronics*: §5.1–8 , 6.5, and appendix 4
Yariv, *Optical Electronics in Modern Communications*: §5.1–3, 6.3–5

Chapter 10

Lasers II: Cavities and examples

In Chapter 9 we pointed out that a laser works by combining an amplifying medium with a resonant cavity. In this chapter we study how the cavity affects the properties of the light emitted by the laser, and then look at a few examples of important lasers, paying particular attention to the mechanism that produces population inversion.

10.1 Laser cavities

The cavity is an essential part of a laser. It provides the positive feedback that turns an amplifier into an oscillator, and determines the properties of the beam of light that is emitted, as shown schematically in Fig. 10.1. This beam is characterized by its **transverse** and **longitudinal** mode structure, which are considered separately below.

10.1.1 Transverse modes

The transverse modes of a laser beam describe the variation of the electrical field across a cross-sectional slice of the beam. The modes are labelled TEM_{mn} where m and n are integers. TEM stands for "transverse electro magnetic". If the field is propagating in the z direction, the (x, y) dependence of the field is given by:

$$\mathcal{E}_{mn}(x,y) = \mathcal{E}_0 H_m\left(\frac{\sqrt{2}x}{w}\right) H_n\left(\frac{\sqrt{2}y}{w}\right) \exp\left(-\frac{x^2+y^2}{w^2}\right) , \tag{10.1}$$

where w is the **beam waist parameter** that determines the size of the beam, and H_m and H_n are mathematical functions called the **Hermite polynomials**.[1] The first few polynomials are:

$$\begin{aligned} H_0(u) &= 1\,, \\ H_1(u) &= 2u\,, \\ H_2(u) &= 2(2u^2-1)\,. \end{aligned} \tag{10.2}$$

The most important mode is the TEM_{00} mode. This has a Gaussian radial distribution:

$$\mathcal{E}_{00}(x,y) = \mathcal{E}_0 \exp\left(-\frac{x^2+y^2}{w^2}\right) = \mathcal{E}_0 \exp\left(-\frac{r^2}{w^2}\right) , \tag{10.3}$$

[1] Hermite polynomials also appear in the solution of the Schrödinger equation for a simple harmonic oscillator.

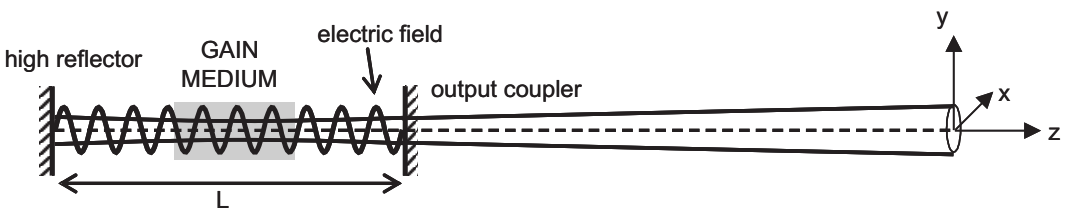

Figure 10.1: Laser cavity and output beam

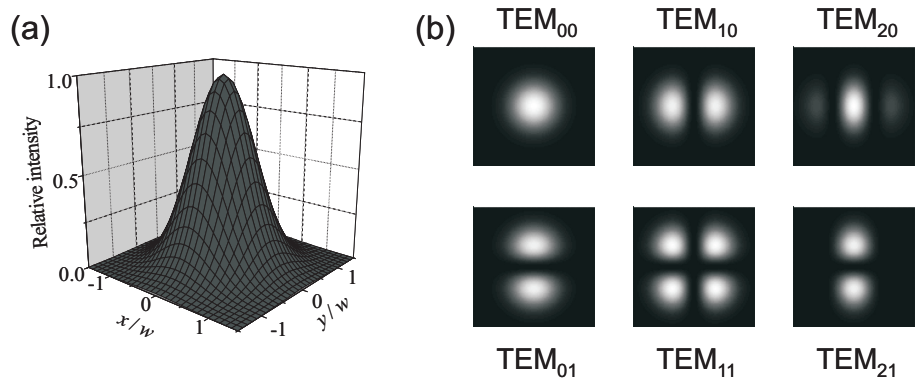

Figure 10.2: (a) Intensity distribution of a TEM_{00} mode, which has a Gaussian profile. (b) Beam profiles produced by various higher-order laser modes. Note that the side lobes in the x direction for the TEM_{21} mode are too faint to be seen on this grey scale.

where r is the distance from the centre of the beam, as shown in Fig. 10.2(a). The TEM_{00} mode is the closest thing to a ray of light found in nature. It has the smallest divergence of all the modes and can be focussed to the smallest size. We therefore usually try hard to prevent the other modes from oscillating. This is achieved by inserting apertures in the cavity which are lossy for the higher-order modes but not for the smaller TEM_{00} mode.

Figure 10.2(b) compares the beam cross-section for a number of higher-order laser modes with that of the TEM_{00} mode. Note that the TEM_{mn} mode has m nodes (zeros) in the x direction and n nodes in the y direction. These higher-order modes make pretty pictures, but are not useful for very much. A well-designed laser will contain apertures that allow only the TEM_{00} mode to oscillate.

10.1.2 Longitudinal modes

The longitudinal modes determine the emission spectrum of the laser. The light bouncing repeatedly off the end mirrors sets up standing waves inside the cavity, as shown in Fig. 10.1. There are nodes (field zeros) at the mirrors because they have high reflectivities. Thus there must be an integer number of half wavelengths inside the cavity. If the length of the cavity is L, this condition can be written:

$$L = \text{integer} \times \frac{\lambda}{2} = \text{integer} \times \frac{c}{2n\nu}\,, \tag{10.4}$$

where n is the average refractive index of the cavity. In gas lasers, n will be very close to unity. It will also be the case that $n \approx 1$ in a solid-state laser with a short laser rod inside a long air-filled cavity. Equation 10.4 implies that only certain frequencies that satisfy:

$$\nu = \text{integer} \times \frac{c}{2nL} \tag{10.5}$$

will oscillate. Most cavities are much larger than the optical wavelength and thus the value of the integer in equations 10.4 and 10.5 is very large.[2] The most important parameter is the longitudinal mode spacing:

$$\Delta\nu_{\text{mode}} = \frac{c}{2nL}\,. \tag{10.6}$$

Table 10.1 lists the longitudinal mode spacing for several lasers.

10.2 Single-mode, multi-mode, and mode-locked lasers

The gain bandwidth of a laser medium will usually be much wider than the spacing of the longitudinal modes of the cavity. This leads to a number of ways of operating the laser.

[2]This is not true for "microcavity lasers", where we deliberately make the cavity to be of similar dimensions to the optical wavelength. In this case the integer would have a value of unity. The use of microcavity semiconductor lasers is now widespread in optical fibre systems.

	Diode laser	HeNe laser	Argon ion laser
L	1 mm	30 cm	2 m
n	3.5	1	1
Mode spacing	150 GHz	500 MHz	75 MHz

Table 10.1: Mode spacing for several common laser

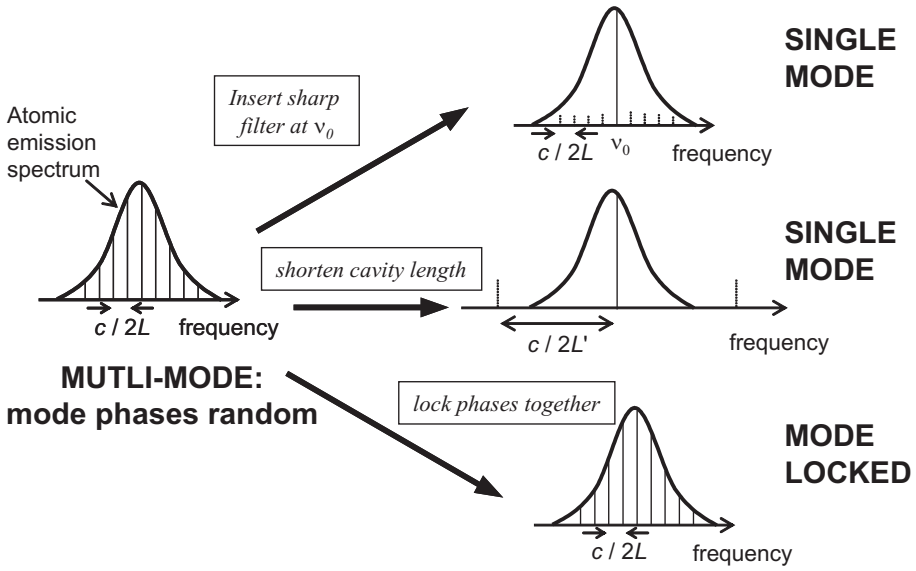

Figure 10.3: Multi-mode, single-mode and mode-locked operation

10.2.1 Multi-mode and single-mode lasers

For a given longitudinal mode to oscillate, its frequency must lie within the emission spectrum of the laser transition. Unless we do something about it, there will be a tendency for all the longitudinal modes that experience gain to oscillate. Therefore the laser will have **multi-mode operation**, as illustrated in Fig. 10.3. As a rough guide, the number of modes that will be oscillating is equal to the gain bandwidth divided by the mode spacing. Thus for a 30 cm HeNe laser with a gain bandwidth of 1.5 GHz, there will be three modes oscillating. In a Doppler-broadened emission line such as that from the Neon atoms in a HeNe laser, the phases of these modes will be random relative to each other because they are emitted by different atoms.

When a laser runs in multi-mode operation, its spectral bandwidth is not significantly smaller than that of the light emitted from the same transition in a discharge lamp. For many applications (e.g. supermarket bar-code readers), this is not an issue. However, for some others, it is. An obvious example is high-resolution spectroscopy. Other examples include those that rely on having high temporal coherence, for example: holography and interferometry. This follows from the fact that the temporal coherence is inversely proportional to the spectral bandwidth. (See Section 10.3 below.) It is therefore interesting to see if we can make the laser run on just a single longitudinal mode. The spectral linewidth would then be determined by the quality factor (Q) of the cavity rather than the gain band width. This is called **single-mode operation**.

One way to achieve to single-mode operation is to shorten the cavity so that the mode spacing exceeds the gain bandwidth. See Fig. 10.3. In this case only one mode will fall within the emission line of the transition and the laser will automatically oscillate on only one mode. However, this may not be practical. For example, in the case of the HeNe laser considered above, we would need to make the cavity shorter than 10 cm. Such a laser would have very small round-trip gain, and we would probably not be able to make it oscillate. A better way to obtain single mode operation is to insert a narrow frequency filter in the cavity such as a Fabry–Perot etalon. By tuning the spacing of the etalon, the frequency of the single mode can be changed continuously, which is very useful for high-resolution spectroscopy. The spectral line width of a single-mode laser is typically a few MHz. This is about a thousand times narrower than the atomic emission line that produces the light.

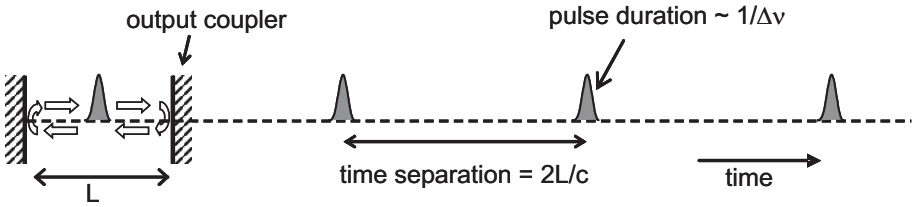

Figure 10.4: Mode-locked laser pulses from a cavity with $n = 1$.

10.2.2 Mode locking

Lasers can be made to operate continuously or in pulses. The length of the pulse might be determined, for example, by the duration of the flash-lamp pulse that produces the population inversion, or by the properties of a Q-switch. (See Section 9.7.) The pulses produced in this way are relatively long, e.g, tens of nanoseconds at best. However, there is another technique called **mode locking** that leads to the emission of a continuous train of very short ("ultrashort") pulses. This is the technique we consider here.

Mode locking is the opposite extreme to single-mode operation. In a mode-locked laser we try to get as many longitudinal modes oscillating as possible, but with all their phases locked together. (See Fig. 10.3). This contrasts with a multi-mode laser in which many modes are oscillating but with random phases with respect to each other.

In Appendix F we prove that the mode-locked operation of a laser corresponds to a single pulse oscillating around the cavity and getting emitted every time it hits the output coupler, as shown in Fig. 10.4. The time taken for a pulse to circulate around a cavity of length L with $n = 1$ is $2L/c$. Therefore we get pulses out of the laser at a repetition rate of $(2L/c)^{-1}$.

The minimum pulse duration is set by the Fourier transform of the gain spectrum:

$$\Delta t_{\min} \Delta\nu \gtrsim 1/2\pi \,, \tag{10.7}$$

where $\Delta\nu$ is the gain bandwidth. This "uncertainty principle" means that to get very short pulses we need a wide gain bandwidth. Gas lasers are not very good in this context because they are based on fairly narrow atomic transitions. For example, the bandwidth of the 632.8 nm line in the HeNe laser is 1.5 GHz (see Table 10.2), and so the pulses that can be produced must be at least 0.11 ns long.

The best results have been achieved in tuneable lasers such as dye lasers or titanium-doped sapphire lasers. The gain bandwidth of the Ti:sapphire laser is nearly 10^{14} Hz, and mode-locked Ti:saphhire lasers routinely produce pulses shorter than 100 fs (1 fs $= 10^{-15}$ s), which corresponds to millions of longitudinal modes oscillating. When the full gain bandwidth of the crystal is used, pulses shorter than 1 fs have been produced from this laser.

Mode locking is achieved by two main techniques. With active mode locking, a time-dependent shutter is inserted in the cavity.[3] The shutter is opened briefly every $2L/c$ seconds. Continuous operation of the laser is impossible, but the mode-locked pulses will be unaffected by the shutter. In passive mode locking, a saturable absorber is inserted in the cavity. Such absorbers have strong absorption at low powers and small absorption at high powers. The peak power in the pulsed mode is much higher than in continuous operation, and thus the cavity naturally selects the pulsed mode.

Mode-locked lasers are widely used in scientific research to study fast processes in physics, chemistry, and biology. For example, the typical time for a current-carrying electron in a copper wire to interact with a phonon at room temperature is about 100 fs. Similarly, the early stages of many chemical reactions or biological processes such as photosynthesis take place in less than 10^{-12} s. Another widespread application of short pulse lasers in biology is in microscopy. It is common practice to obtain images of biological molecules by tagging them with fluorescent chromophores (e.g. dyes, quantum dots) and then exciting the sample with a laser in a confocal microscopy. The use of mode-locked lasers gives far superior depth resolution compared to continuous wave (CW) lasers.[4]

Mode-locked lasers are also useful to telecommunication companies, who are interested in packing as many bits of information (represented by pulses of light) as possible down their optical fibres. The shorter the pulses, the higher the data rate. There are also medical applications: it is much cleaner to

[3] The time-dependent shutter is typically made by using a high speed acousto-optic modulator.

[4] When a mode-locked laser is used, the higher peak power allows the fluorescent chromophore to be excited by two-photon absorption. The power is only high enough for this to occur at the focus of the laser, and so only the part of the sample at the focus produces light. With a CW laser, by contrast, the chromophores are excited by standard one-photon absorption, and the whole depth of the sample emits light.

SOURCE	Spectral line width $\Delta\nu$ (Hz)	Coherence time t_c (s)	Coherence length l_c
Sodium discharge lamp (D-lines at 589 nm)	5×10^{11}	2×10^{-12}	0.6 mm
Multi-mode HeNe laser (632.8nm line)	1.5×10^{9}	6×10^{-10}	20 cm
Single-mode HeNe laser (632.8 nm line)	1×10^{6}	10^{-6}	300 m

Table 10.2: Coherence length of several light sources.

use a very short, low-energy, high-peak-power pulse for laser surgery, than a longer pulse with the same peak power but much higher energy.

10.3 Coherence of laser light

As mentioned in Section 9.1, laser light has a high degree of both spatial and temporal coherence. The spatial coherence is related to the phase uniformity across a cross-sectional slice of the beam. When the laser is running in a well-defined transverse mode, the optical phase across such a slice will be constant. Hence the spatial coherence follows from the transverse modes, and will be very high when the laser is running on a single transverse mode.

The temporal coherence of light refers to the time duration over which the phase is constant. In general, the temporal coherence time t_c is determined by the spectral line width $\Delta\nu$ according to:

$$t_c \sim \frac{1}{\Delta\nu}\,. \tag{10.8}$$

Hence the coherence length l_c is given by:

$$l_c = ct_c \sim \frac{c}{\Delta\nu} \tag{10.9}$$

Typical values of the coherence length for a number of light sources are given in Table 10.2. The figures explain why it is much easier to do interference experiments with a laser than with a discharge lamp. If the path difference exceeds l_c you will not get interference fringes, because the light is incoherent. In the case of the single-mode HeNe laser, you can set up an interferometer in which the path lengths differ by 300 m, and you will still observe fringes. The long coherence length of laser light is useful in holography and interferometry.

10.4 Examples of lasers

There are many different types of lasers in common use, and it is not possible to describe all of them here. Most lasers operate at fixed wavelengths:

Infrared lasers CO_2 (10.6 μm), erbium (1.55 μm), Nd:YAG (1.064 μm), Nd:glass (1.054 μm);

Visible lasers ruby (693 nm), krypton ion (676, 647 nm), HeNe (633 nm), copper vapour (578 nm), doubled Nd:YAG (532 nm), argon ion (514, 488 nm), HeCd (442 nm);

Ultraviolet lasers argon ion (364, 351 nm), tripled Nd:YAG (355 nm), nitrogen (337 nm), HeCd (325 nm), quadrupled Nd:YAG (266 nm), excimer (308, 248, 193, 150 nm).

Others lasers are tuneable, for example: dye lasers (typical tuning range ∼ 100 nm, dyes available from UV to near infrared); Ti:sapphire lasers (700-1000 nm, doubled: 350-500nm); free electron lasers (far infrared to ultraviolet). The most common lasers in widespread use are semiconductor diode lasers. Cheap and efficient diode lasers available at blue (400 nm), red (620-670 nm), and near-infrared wavelengths (700-1600 nm).

In the sections below we consider a few of the more important lasers that are available, following the general classification according to whether the gain medium is a gas or a solid.[5]

[5]There are relatively few liquid-phase lasers. The most important examples are dye lasers. However, with the advent of broadly-tuneable high power solid-state lasers such as Ti:sapphire lasers, and the development of techniques of nonlinear optics to extend their frequency range (see Appendix G), dye lasers are gradually becoming obsolete.

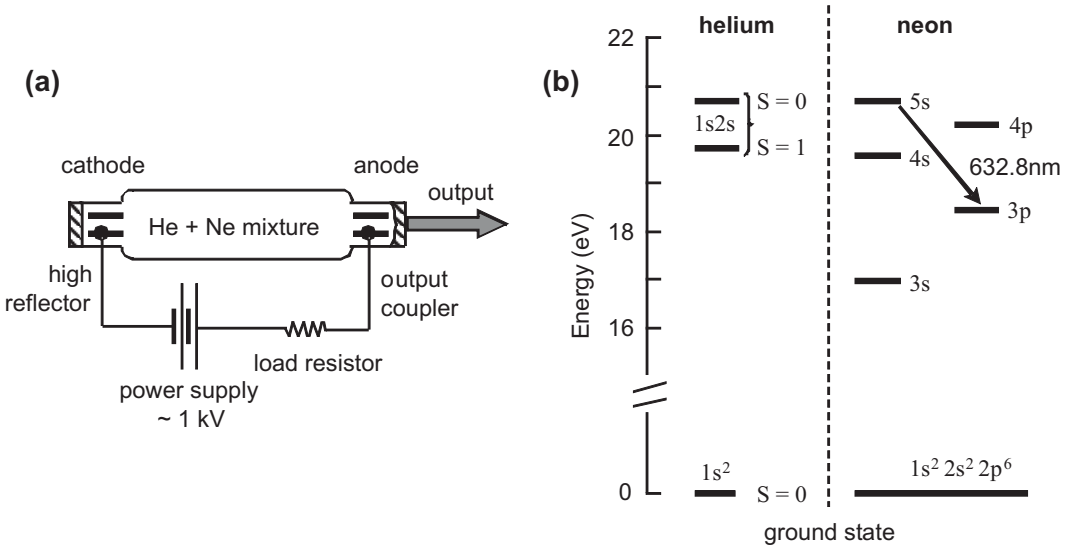

Figure 10.5: (a) Schematic diagram of a HeNe laser. (b) Level scheme for the HeNe laser.

10.5 Gas lasers

10.5.1 The helium-neon (HeNe) laser

Helium-neon lasers consist of a discharge tube inserted between highly reflecting mirrors, as shown schematically in Fig. 10.5(a). The tube contains a mixture of helium and neon atoms in the approximate ratio of He:Ne 5:1. By applying a high voltage across the tube, an electrical discharge can be induced. The electrons collide with the atoms and put them into excited states. The light is emitted by the neon atoms, and the purpose of the helium is to assist the population inversion process. To see how this works we need to refer to the level diagram in Fig. 10.5(b).

Helium has two electrons. In the ground state both electrons are in the 1s level. The first excited state is the 1s2s configuration. There are two possible energies for this state because there are two possible configurations of the electron spin: the singlet $S = 0$ and the triplet $S = 1$ terms. The helium atoms are excited by collisions with the electrons in the discharge tube and cascade down the levels. When they get to the 1s2s configuration, however, the cascade process slows right down. In the 2s1s $\rightarrow 1s^2$ transition one of the electrons jumps from the 2s level to the 1s level. This is forbidden by the $\Delta l = \pm 1$ selection rule. Furthermore, transitions from the 1s2s $S = 1$ level to the $1s^2$ $S = 0$ ground state are also forbidden by the $\Delta S = 0$ selection rule. The net result is that all transitions from the 1s2s levels are strongly forbidden. The 1s2s level therefore has a very long lifetime, and is called **metastable**. See Section 6.5 in Chapter 6 for more details.

Neon has ten electrons in the configuration $1s^2 2s^2 2p^6$. The excited states correspond to the promotion of one of the 2p electrons to higher levels. This gives the level scheme shown in the diagram. The symbols of the excited states refer to the level of this single excited electron. By good luck, the 5s and 4s levels of the neon atoms are almost degenerate with the $S = 0$ and $S = 1$ terms of the 1s2s configuration of helium. Thus the helium atoms can easily de-excite by collisions with neon atoms in the ground state according to the following scheme:

$$\mathrm{He}^* + \mathrm{Ne} \Rightarrow \mathrm{He} + \mathrm{Ne}^* \,. \tag{10.10}$$

The star indicates that the atom is an excited state. Any small differences in the energy between the excited states of the two atoms are taken up as kinetic energy.

This scheme leads to a large population of neon atoms in the 5s and 4s excited states. This gives population inversion with respect to the 3p and 4p levels. It would not be easy to get this population inversion without the helium because collisions between the neon atoms and the electrons in the tube would tend to excite all the levels of the neon atoms equally. This is why there is more helium than neon in the tube.

The main laser transition at 632.8 nm occurs between the 5s level and the 3p level. The lifetime of the 5s level is 170 ns, while that of the 3p level is 10 ns. This transition therefore easily satisfies the criterion $\tau^{\mathrm{upper}} > \tau^{\mathrm{lower}}$. (See discussion of eqn 9.28.) This ensures that atoms do not pile up in the

lower level once they have emitted the laser photons, as this would destroy the population inversion. The atoms in the 3p level rapidly relax to the ground state by radiative transitions to the 3s level and then by collisional de-excitation to the original 2p level. Lasing can also be obtained on other transitions: for example, 5s $\rightarrow$ 4p at 3391 nm and 4s $\rightarrow$ 3p at 1152 nm. These are not as strong as the main 632.8 nm line.

The gain in a HeNe tube tends to be rather low because of the relatively low density of atoms in the gas (compared to a solid). This is partly compensated by the fairly short lifetime of 170 ns. (See eqn 9.24.) The round trip gain may only be a few percent, and so very highly reflecting mirrors are needed. With relatively small gain, the output powers are not very high - only a few mW. However, the ease of manufacture makes these lasers to be extremely common for low power applications: bar-code readers, laser alignment tools (theodolites, rifle sights), classroom demos etc. They are gradually being replaced nowadays by visible semiconductor laser diodes, which are commonly used in laser pointers.

10.5.2 Helium-cadmium lasers

The HeCd laser is another gas laser system based on helium. The population inversion scheme in HeCd is similar to that in HeNe except that the active medium is Cd^+ ions. The laser transitions occur in the blue and the ultraviolet at 442 nm, 354 nm and 325 nm. The UV lines are useful for applications that require short wavelength lasers, such as high-precision printing on photosensitive materials. Examples include lithography of electronic circuitry and making master copies of compact disks.

10.5.3 Ion lasers

There are several important types of gas lasers that use ions rather than neutral atoms as the gain medium, for example, the **argon-ion laser**. The argon ions are produced by collisions with electrons in a discharge tube. The atomic number of argon is 18, and so the Ar^+ ion has 17 electrons in the configuration $1s^2 2s^2 2p^6 3s^2 3p^5$. The excited states of the Ar^+ ion are generated by exciting one of the five 3p electrons to higher levels, and the most important laser transitions occur between the 4p and 4s levels. Spin-orbit coupling splits this into a doublet, with emission lines at 488 nm (blue) and 514.5 nm (green). The **krypton ion laser** works by similar principles, and has a strong laser emission line in the red at 676.4 nm. This red line can be combined with the green and blue lines of the argon-ion laser to make very colourful laser light shows.

In addition to laser light shows, argon-ion lasers are frequently used for pumping tuneable lasers such as dye lasers and Ti:sapphire lasers. There are also some medical applications such as laser surgery, and scientific applications include fluorescence excitation and Raman spectroscopy.

10.5.4 Carbon dioxide lasers

The CO_2 laser is one of the best examples of a molecular laser. The transitions take place between the vibrational levels of the molecule. The strongest emission lines are in the infrared around 10.6 μm. The lasers are very powerful with powers up to several kilowatts possible. Hence they are used in cutting applications in industry (including the military industry!) and also for medical surgery. The high power output is a consequence of the fact that the stimulated emission becomes more favourable compared to spontaneous emission at lower frequencies: see eqn 9.14.

A mixture of nitrogen and CO_2 in a ratio of about 4:1 is used in the laser tube. The N_2 molecules are excited by collisions with electrons, and then transfer their energy to the upper level of the CO_2 molecules. This gives population inversion in much the same sort of way as for the HeNe mixture.

10.6 Solid-state lasers

10.6.1 Ruby lasers

Ruby lasers have historical importance because they were the first successful laser to operate. Ruby consists of Cr^{3+} ions doped into crystalline Al_2O_3 (sapphire) at a typical concentration of around 0.05% by weight. The Al_2O_3 host crystal is colourless. The light is emitted by transitions of the Cr^{3+} impurities.

The level scheme for ruby is shown in Fig. 10.6(a). Ruby is a three-level system (see Section 9.8), with strong absorption bands in the blue and green spectral regions. (Hence the red colour: *ruber* means "red" in Latin.) Electrons are excited to these bands by a powerful flashlamp. These electrons relax rapidly to the upper laser level by non-radiative transitions in which phonons are emitted. This leads

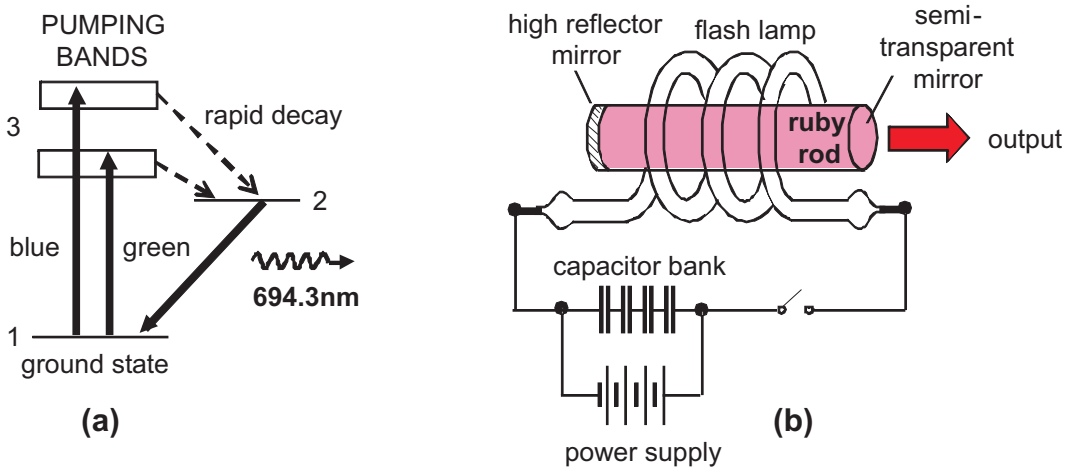

Figure 10.6: (a) Level diagram for ruby (Cr^{3+}:Al_2O_3). (b) Schematic diagram of a ruby laser.

to a large population in the upper laser level. If the flashlamp is powerful enough, it will be possible to pump more than half of the atoms from the ground state (level 1) to the upper laser level (level 2). In this case, there will then be population inversion between level 2 and level 1, and lasing can occur if a suitable cavity is provided. The laser emission is in the red at 694.3 nm.

Figure 10.6(b) shows a typical arrangement for a ruby laser. The crystal is inserted inside a powerful helical flashlamp. Water-cooling prevents damage to the crystal by the intense heat generated by the lamp. Mirrors at either end of the crystal define the cavity. Reflective coatings can be applied directly to the end of the rod as shown, or external mirrors can be used (not shown). The lamps are usually driven in pulsed mode by discharge from a capacitor bank. The pulse energy can be as high as 100 J per pulse. This is because the upper laser level has a very long lifetime (3 ms) and can store a lot of energy.

10.6.2 Neodymium lasers (Nd:YAG and Nd:glass)

Neodymium ions form the basis for a series of high power solid-state lasers. In the two most common variants, the Nd^{3+} ions are doped into either Yttrium Aluminium Garnet (YAG) crystals or into a phosphate glass host. These two lasers are known as either Nd:YAG or Nd:glass. The main laser transition is in the near-infrared at about 1.06 μm. The wavelength does not change much on varying the host.

Figure 10.7(a) shows the level scheme for the Nd^{3+} lasers, which are four-level lasers. Electrons are excited to the pump bands by absorption of photons from a powerful flashlamp or from a diode laser operating around 800 nm. The electrons rapidly relax to the upper laser level by phonon emission. Lasing then occurs on the $^4F_{3/2} \rightarrow {}^4 I_{11/2}$ transition.[6] The electrons return to the ground state by rapid non-radiative decay by phonon emission.

Figure 10.7(b) shows the cavity arrangement in a flashlamp-pumped system. The rod and lamp are positioned at the foci of an elliptical reflector. This ensures that most of the photons emitted by the lamp are incident on the rod to maximize the pumping efficiency. Mirrors at either end of the rod provide the optical cavity. The laser can either be operated in pulsed or continuous wave mode.

As with the ruby laser, the lifetime of the upper laser level is long: 0.2–0.3 ms, depending on the host. This long lifetime, which is a consequence of the fact that the laser transition is E1-forbidden, allows the storage of large amounts of energy. Continuous wave Nd:YAG lasers can easily give 20–30 W, while pulsed versions can give energies up to 1 J in 10 ns. The pulse energies possible from Nd:glass lasers are even higher, although they can only operate at lower repetition rates. The Lawrence Livermore Lab in California uses Nd:glass lasers for fusion research. The pulse energy in these systems is ~ 10 kJ. With pulse durations in the 10 ns range, this gives peak powers of 10^{12} W.

Nd lasers are extensively used in industry for cutting applications, and in medicine for laser surgery. They are very rugged and can be used in extreme conditions (eg onboard military aircraft). Frequency-doubled Nd:YAG lasers (see Appendix G) are now gradually replacing argon-ion lasers for pumping

[6] This transition is strongly forbidden for free atoms. However, the wave functions of the Nd^{3+} ion get distorted in the crystal by the electric fields from the neighbouring host atoms, and this relaxes the selection rules. The Einstein A coefficient of the 1064 nm transition in Nd:YAG is $4.3 \times 10^3\ s^{-1}$.

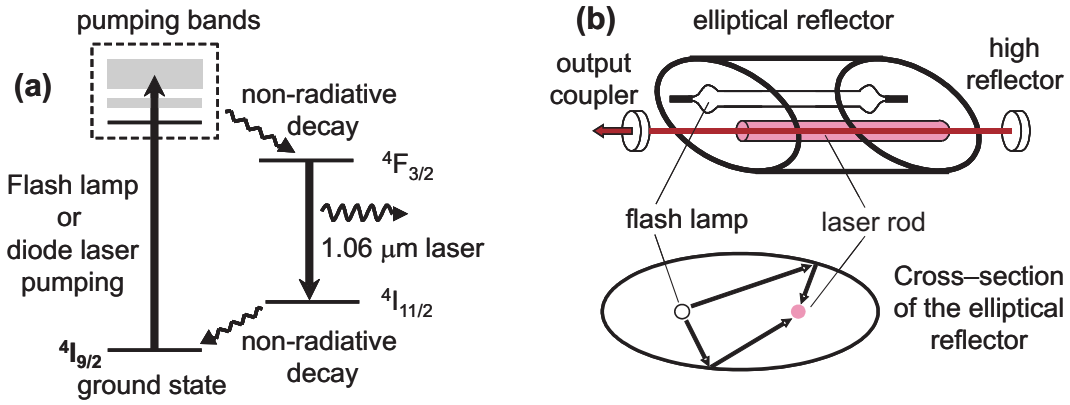

Figure 10.7: (a) Level diagram for the Nd^{3+} lasers. (b) Schematic diagram of an Nd:YAG laser.

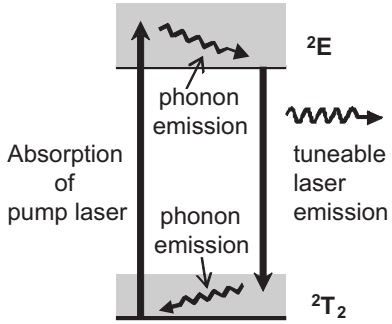

Figure 10.8: Level diagram for the Ti:sapphire laser.

tuneable lasers such as Ti:sapphire. (See below)

10.6.3 Ti:sapphire

Titanium-doped sapphire lasers represent the current state-of-the-art in tuneable lasers. The level scheme is shown in Fig. 10.8. The active transitions occur in the Ti^{3+} ion. This has one electron in the 3d shell. In the octahedral environment of the sapphire (Al_2O_3) host, the crystal field splits the five m levels of the 3d shell into a doublet and a triplet. These are labelled as the ^2E and $^2T_2$ states in Fig. 10.8.[7] The electron-phonon coupling in Ti:sapphire is very strong, and the ^2E and $^2T_2$ states are broadened into "vibronic" bands. The absorption of the ^2E band peaks in the green-blue spectral region, and thus can be pumped by the 488 nm and 514 nm lines of an argon ion laser. Alternatively, a frequency-doubled Nd:YAG laser operating at ($1064/2 = 532$ nm) can be used.

Electrons excited into the middle of the ^2E band rapidly relax by phonon emission to the bottom of the band. Laser emission can then take place to anywhere in the $^2T_2$ band. The electrons finally relax to the bottom of the $^2T_2$ band by rapid phonon emission.

The fact that tuning can be obtained over the entire $^2T_2$ band is a very useful feature because it means that the laser wavelength can be chosen at will. Lasing has in fact been demonstrated all the way from 690 nm to 1080 nm i.e. over nearly 400 nm. This is why it makes sense to use one laser to pump another: we convert a fixed frequency laser such as the argon-ion or frequency-doubled Nd:YAG into a tuneable source. Energy conversion efficiencies of up to 25% are possible.

The broad emission band width is also ideal for making short pulse mode-locked lasers. (See Section 10.2.2). The shortest pulses that can be produced are given by $\Delta t \Delta\nu \gtrsim 1/2\pi$. With such a broad emission band, it has been possible to generate pulses shorter than 10 fs.

[7] This notation might be familiar to the Chemical Physicists. The letters are abbreviations for German words. "E" and "T" label doublet and triplet states. The superscript of 2 refers to the spin degeneracy. Thus these two states contain $(2 \times 2 + 2 \times 3 = 10)$ levels, as we would expect for the 3d states. The subscript of 2 on the triplet state indicates that it has a particular symmetry.

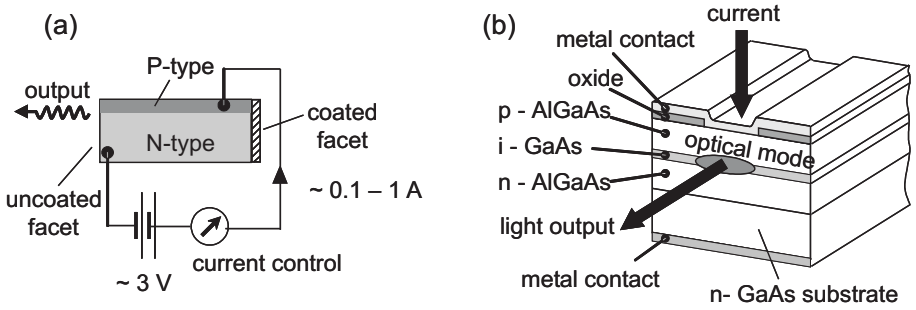

Figure 10.9: (a) Schematic diagram of the operation of a semiconductor diode laser. (b) Detailed sketch of a typical GaAs diode laser chip.

10.6.4 Semiconductor diode lasers

Semiconductor diode lasers are by far the most common types of lasers. They are used in laser printers, DVD players, laser pointers, and optical fibre communication systems. The laser consists of a semiconductor p-n diode cleaved into a small chip, as shown in Fig. 10.9(a). Electrons are injected into the n-region, and holes into the p-region. At the junction between the n- and p-regions we have both electrons in the conduction band and holes (i.e. empty states) in the valence band. This creates population inversion between the conduction and valence bands, and gain is produced at the band gap energy E_g of the semiconductor. The electrons in the conduction band drop to the empty states in the valence band, and laser photons with energy $h\nu = E_g$ are emitted. The drive voltage must be at least equal to E_g/e, where e is the electron charge.

The laser cavity is formed by using the cleaved facets of the chips. The refractive index of a typical semiconductor is in the range 3-4, which gives about 30% reflectivity at each facet. This is enough to support lasing, even in crystals as short as $\sim 1\,\mathrm{mm}$, because the gain in the semiconductor crystal is very high. A highly reflective coating is often placed on the rear facet to prevent unwanted losses through this facet and hence reduce the threshold.

The semiconductor must have a *direct* band gap to be an efficient light emitter. Silicon has an indirect band gap, and is therefore not used for laser diode applications. The laser diode industry is based mainly on the compound semiconductor GaAs, which has a direct band gap at 1.4 eV (890 nm). A typical design of a GaAs diode laser is shown in Fig. 10.9(b). By using alloys of GaAs, the band gap can be shifted into the red spectral region for making laser pointers, or further into the infrared to match the wavelength for lowest losses in optical fibres (1500 nm). Blue laser diodes are made from the wide band gap III–V semiconductor GaN and its alloys. These lasers are used in "blue-ray" systems.

The power conversion efficiency of electricity into light in a diode laser is very high, with figures of 25% typically achieved. This compares with typical efficiencies of $< 0.1\%$ in gas lasers. Since the laser chips are so small, it is possible to make high power diode lasers by running many GaAs chips in parallel. Laser power outputs over 20W can easily be achieved in this way. These high power laser diodes are very useful for pumping Nd:YAG lasers.

Reading

Hooker and Webb, *Laser Physics*: §6.3, chapter 7, §8.3, chapters 9, 11
Wilson and Hawkes, *Optoelectronics*: §5.9, 5.10.1–3, 6.1–3
Bransden and Joachain, *Physics of Atoms and Molecules*, §15.1
Demtröder, *Atoms, Molecules and Photons*: §8.2–4, 8.6
Smith and King, *Optics and Photonics*, §15.2, 15.7, 15.9, chapter 16, §17.4
Hecht, *Optics*: §13.1
Yariv, *Optical Electronics in Modern Communications*: §6.6–7, chapter 7
Silfvast, *Laser Fundamentals*, §2.4, chapters 10–15
Svelto, *Principles of Lasers*: §7.7-8, 8.6, chapters 9–10

Chapter 11

Laser cooling of atoms

11.1 Introduction

The resonant force between an atom and a light field was first observed in 1933, when Frisch measured the deflection of a sodium beam caused by a sodium lamp shining on the side of the beam. The invention of lasers opened up new possibilities, and the first **laser cooling** experiments were carried out in the 1980s. The importance of this work was recognized by the award of the Nobel Prize for Physics in 1997 to three of the pioneers of the field: Stephen Chu, Claude Cohen-Tannoudji, and William D. Phillips.

There are two aspects of laser cooling that make it particularly remarkable.

1. It is highly surprising that the technique works at all. We would normally expect a powerful laser to cause heating rather than cooling. This makes us realize that the technique will only work when special conditions are satisfied. These will be discussed in the rest of this chapter.

2. The very low temperatures achieved by laser cooling are extremely impressive, but this in itself is not the main point. Techniques for achieving very low temperatures have been used for decades by condensed matter physicists. For example, commercial **dilution refrigerators** routinely achieve temperatures in the milli-Kelvin range, and as early as the 1950s, Nicholas Kurti and co-workers at Oxford University used **adiabatic demagnetisation** to achieve nuclear spin temperatures in the micro-Kelvin range. The novelty of laser cooling is that it produces of an ultracold *gas* of atoms, in contrast to the condensed matter techniques which all work on liquids or solids. These ultracold atoms only interact weakly with each other, which makes it possible to study the light-matter interaction with unsurpassed precision.

These aspects of laser cooling have given rise to a whole host of related benefits. Atomic clocks have been made with ever greater accuracy, and a whole range of new quantum phenomena have been discovered. The most spectacular of these is **Bose–Einstein condensation**, which was observed for the first time in 1995.

11.2 Gas temperatures

In order to understand how laser cooling works, we first need to clarify how the temperature of a gas of atoms is measured. The key point is the link between the thermal motion of the atoms and the temperature. Starting from the Maxwell–Boltzmann distribution (cf. eqn 3.40), it is possible to define a number of different characteristic velocities for the gas. The simplest of these is the root-mean-square (rms) velocity, which can be evaluated by remembering the principle of **equipartition of energy**. This states that the average thermal energy per degree of freedom is equal to $\frac{1}{2}k_{\rm B}T$. For an atom of mass m, each component of the velocity must therefore satisfy:

$$\frac{1}{2}m\overline{v}_i^2 = \frac{1}{2}k_{\rm B}T\,, \tag{11.1}$$

which implies that the rms velocity is given by:

$$\frac{1}{2}m(v^{\rm rms})^2 = \frac{3}{2}k_{\rm B}T\,. \tag{11.2}$$

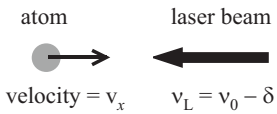

Figure 11.1: In Doppler cooling, the laser frequency is tuned below the atomic resonance by δ. The frequency seen by an atom moving towards the laser is Doppler shifted up by $\nu_0(v_x/c)$.

We therefore conclude that:[1]

$$v_x^{\rm rms} = \sqrt{\frac{k_{\rm B}T}{m}}\,, \tag{11.3}$$

$$v^{\rm rms} = \sqrt{\frac{3k_{\rm B}T}{m}}\,. \tag{11.4}$$

These simple relationships allow us to work out, for example, that the atoms in a typical gas at room temperature jostle about in a random way with thermal velocities of around 1000 kmph. The random thermal motion is the cause of the Doppler broadening of spectral lines considered in Section 3.10.

The link between temperature and the velocity distribution tells us that we can cool the gas if we can slow the atoms down, which is the strategy adopted in laser cooling experiments. Furthermore, the temperature of the gas can be inferred from a measurement of the velocity distribution of the atoms. This is the method that is used to determine the temperature of an ultra cold gas cooled by a laser.

11.3 Doppler Cooling

11.3.1 The laser cooling process

Consider an atom emitting at ν_0 moving in the $+x$ direction towards a laser of frequency $\nu_{\rm L}$ with velocity v_x as shown in Fig. 11.1. The laser is tuned so that its frequency is below the resonance line by an amount δ:

$$\nu_{\rm L} = \nu_0 - \delta\,. \tag{11.5}$$

The Doppler-shifted frequency $\nu_{\rm L}^{\rm observed}$ of the laser in the atom's frame of reference is given by:

$$\nu_{\rm L}^{\rm observed} = \nu_{\rm L}\left(1+\frac{v_x}{c}\right) = (\nu_0-\delta)\left(1+\frac{v_x}{c}\right) = \nu_0 - \delta + \frac{v_x}{c}\nu_0 - \frac{v_x}{c}\delta\,. \tag{11.6}$$

The last term is small because $\delta \ll \nu_0$ and $v_x \ll c$. Hence if we choose

$$\delta = \nu_0\frac{v_x}{c}\,, \tag{11.7}$$

we find $\nu_{\rm L}^{\rm observed} = \nu_0$. This situation is depicted in Fig. 11.2(a). The laser is in resonance with atoms moving in the $+x$ direction, but not with those moving away or obliquely. For sodium at 300 K with $v_x \sim 330$ ms^{-1}, we need to choose $\delta = 560$ MHz for the D-lines at 589 nm. This means that only those atoms moving towards the laser absorb photons from the laser beam.

Now consider what happens after the atom has absorbed a photon from the laser beam. The atom goes into an excited state and then re-emits another photon by spontaneous emission. This occurs on average after a time τ (the radiative lifetime), and the direction of the emitted photon is random. The absorption-emission cycle is illustrated schematically in Fig. 11.2(b).

Repeated absorption-emission cycles generate a net force in the same direction as the laser beam, that is, the $-x$ direction. This happens because each photon of wavelength λ has a momentum of h/λ. Conservation of momentum demands that every time a photon is absorbed from the laser beam the momentum of the atom changes by $(-h/\lambda)$. On the other hand, the change of momentum due to the recoil of the atom after spontaneous emission averages to zero, because the photons are emitted in random directions. Hence the net change of momentum per absorption-emission cycle is given by:

$$\Delta p_x = -\frac{h}{\lambda}\,. \tag{11.8}$$

[1]The rms velocity of the atoms of a beam of atoms effusing from a hot oven differs from eqn 11.3 by factor of 2, and the most probable velocity by factor of $\sqrt{3}$. These numerical factors arise from the fact that the probability of escaping from the oven is related to the velocity, which modifies the probability distribution of the atoms in the beam. These finer details needs not concern us here.

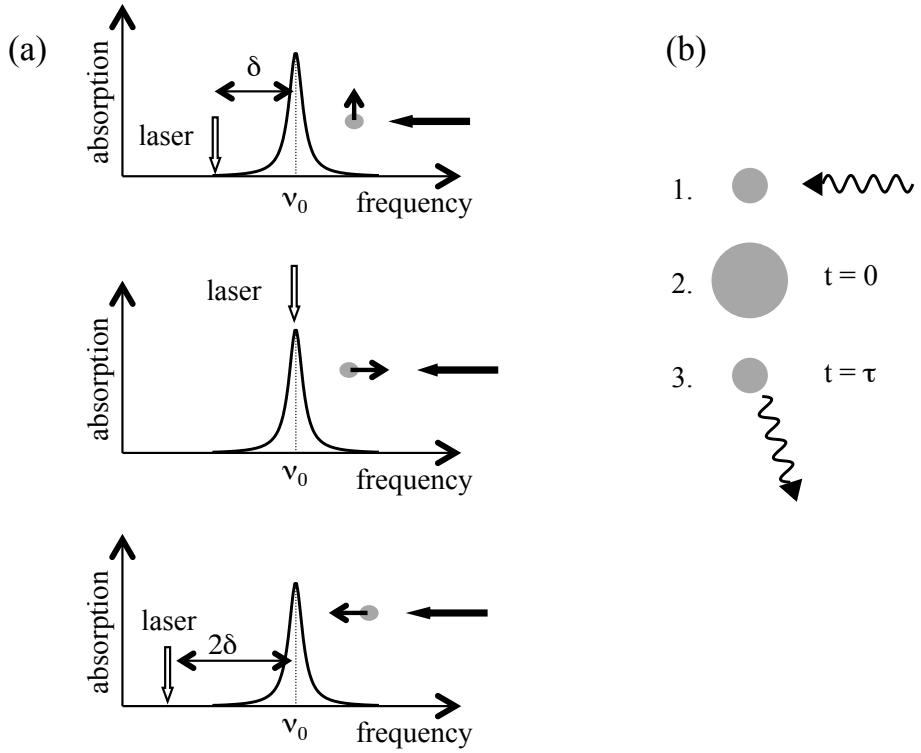

Figure 11.2: Doppler cooling. (a) Doppler-shifted laser frequency in the rest frame of the atom. A laser with frequency $\nu_0 - \delta$ is in resonance with the atoms when they are moving towards the laser, but not if they are moving sideways or away, if $\delta = \nu_0(v_x/c)$. (b) An absorption-emission cycle. (1) A laser photon impinges on the atom. (2) The atom absorbs the photon and goes into an excited state. (3) The atom re-emits a photon in a random direction by spontaneous emission after a time τ.

If the laser intensity is large, then the probability for absorption will be large, and the absorption process will be fast. Hence the time to complete the absorption-emission cycle is determined by the radiative lifetime τ. The maximum force exerted on the atom is thus given by:

$$F_x = \frac{\mathrm{d}p}{\mathrm{d}t} = \frac{\Delta p_x}{\tau} = -\frac{h}{\lambda\tau}\,, \tag{11.9}$$

and the deceleration is given by

$$\dot{v}_x = \frac{F_x}{m} = -\frac{h}{m\lambda\tau}\,. \tag{11.10}$$

For the sodium D-lines with $\lambda = 589\,\mathrm{nm}$ and $\tau = 16\,\mathrm{ns}$, we find $F_x = -7.0 \times 10^{-20}\,\mathrm{N}$ and $\dot{v}_x = -1.8 \times 10^6\,\mathrm{ms}^{-2} \sim 10^5 g$.

The number of absorption-emission cycles required to stop the atom is given by:

$$\mathsf{N}_{\mathrm{stop}} = \frac{mu_x}{\Delta p_x} = \frac{mu_x\lambda}{h}\,, \tag{11.11}$$

where u_x is the initial velocity of the atom. This sets a minimum time for the laser beam to slow the atoms to a halt:

$$t_{\mathrm{min}} = \mathsf{N}_{\mathrm{stop}} \times \tau = \frac{mu_x\lambda\tau}{h}\,. \tag{11.12}$$

In this time, the atoms move a minimum distance d_{min} given by:

$$0 - u_x^2 = 2\,\dot{v}_x\,d_{\mathrm{min}}\,, \tag{11.13}$$

where $\dot{v}_x$ is the deceleration given by eqn 11.10, and we have assumed that the final velocity of the atom is very small. This gives:

$$d_{\mathrm{min}} = -\frac{u_x^2}{2\dot{v}_x} = \frac{m\lambda\tau u_x^2}{2h}\,. \tag{11.14}$$

For our standard sodium example with $u_x = 330\,\mathrm{ms}^{-1}$, we find $\mathsf{N}_{\mathrm{stop}} = 1.1 \times 10^4$, $t_{\mathrm{min}} = 0.18$ ms and $d_{\mathrm{min}} = 0.03$ m.

The analysis above ignores stimulated emission. The atom in the excited state — step 2 in Fig. 11.2(b) — can be triggered to emit a photon by stimulated emission from other impinging laser photons. The stimulated photon will be emitted in the same direction as the incident photon, and the photon recoil exactly cancels the momentum kick given by the absorption process. When stimulated emission is included in the analysis, the force is reduced by a factor two at high laser powers. This happens because the population of levels 1 and 2 equalize at a value of $N_0/2$, where N_0 is the total number of atoms. The final result is that the time to stop the atoms and the distance travelled in that time are both doubled.

11.3.2 The Doppler limit temperature

At first sight, we might think that we would be able to completely stop the atoms by the Doppler cooling technique. However, the minimum temperature that can be achieved is set by the uncertainty principle. The cooling effect only works if we have the right detuning frequency δ for the particular velocity. However, from eqn 3.33 we see that the radiative lifetime τ of the transition causes broadening. This gives rise to an intrinsic uncertainty in the energy of the atom, and we will therefore never be able to reduce the thermal energy below:

$$E_{\mathrm{min}} \sim \frac{1}{2}h\Delta\nu_{\mathrm{lifetime}} = \frac{1}{2}h\frac{1}{2\pi\tau} = \frac{\hbar}{2\tau}\,. \tag{11.15}$$

On equating E_{min} with $k_{\mathrm{B}}T_{\mathrm{min}}$, we then find:

$$T_{\mathrm{min}} = \frac{\hbar}{2k_{\mathrm{B}}\tau}\,. \tag{11.16}$$

This minimum temperature is called the **Doppler limit**. The equivalent minimum speed is found by setting E_{min} equal to $\frac{1}{2}mv_{\mathrm{min}}^2$. For the sodium D-lines with $\tau = 16\,\mathrm{ns}$, the Doppler limit temperature is $2.4 \times 10^{-4}\,\mathrm{K} \equiv 240\,\mu\mathrm{K}$, and the minimum thermal velocity is around $0.4\,\mathrm{ms}^{-1}$.

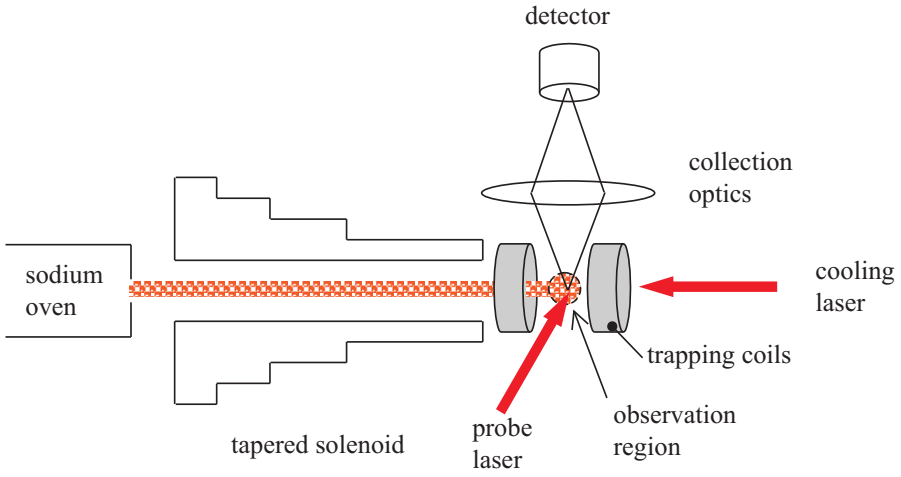

Figure 11.3: William D. Phillips' apparatus at the NIST laboratory, USA, to stop a beam of sodium atoms. The frequency of the cooling laser is fixed, and the transition energy of the atoms is shifted in a controlled way by the Zeeman effect using a tapered solenoid. The trapping coils prevent the atoms falling under gravity. A probe laser is used to measure the velocity distribution of the cooled atoms.

11.4 Experimental considerations

Efficient cooling of the atoms requires that the laser should exert the optimal force on the atoms, which occurs when the laser is detuned by the amount set out in eqn 11.7. However, the velocity of the atoms decreases as the atoms cool, which suggests two possible strategies to achieve low temperatures:

1. Tune the laser frequency in a programmed way as the atoms slow down.

2. Keep the laser frequency fixed and tune the transition frequency.

The first method is called "chirp" cooling, in analogy to the chirping sound made by birds, in which the frequency of the sound changes during the birdsong. Early experiments used tunable dye lasers, but more modern experiments on cesium use tunable semiconductor diode lasers.

An ingenious approach that follows the second method is shown in Fig. 11.3. The sodium atoms are produced by heating sodium metal in an oven to 450°C. Some of the atoms emerge through a small hole in the oven, and are collimated into a beam moving in the $+x$ direction by a second aperture about 10 cm away from the source. The atoms pass through a carefully-designed tapered solenoid, which shifts the the transition energy by the Zeeman effect (see eqn 8.17):

$$h\nu(x) = h\nu_0 + g_J \mu_B B(x) M_J \,, \tag{11.17}$$

where $B(x)$ is the spatially–varying magnetic field. For the $^2S_{1/2}$ ground state of the 3s electron in sodium, we have pure spin angular momentum, and hence $g_J = 2$ and $M_J = \pm 1/2$. If the laser is tuned to ν_0, then the laser cooling condition given in eqn 11.7 is satisfied for the $M_J = +1/2$ state when $\mu_B B(x) = h\nu_0(v_x/c)$. The solenoid was therefore designed so that the reduction of the field strength compensates for the reduction of the velocity as the atoms slow down due to the laser cooling process. Coils were added at the end to prevent the ultra cold atoms falling out of the apparatus under gravity. The atoms with $M_J = -1/2$ are not cooled and escape from the apparatus.

The properties of the cooled atoms can be measured by a second laser. This excites fluorescence which is collected and imaged onto a detector. The velocity distribution can be measured by the "time-of-flight" technique in which we turn the cooling laser off and then watch the gas of trapped atoms expand as a function of time.

11.5 Optical molasses and magneto-optical traps

The arrangement with a single laser beam shown in Fig. 11.1 is able to stop the atoms moving in the positive direction for one of the components of the velocity (i.e the $+x$ direction). To stop the atoms

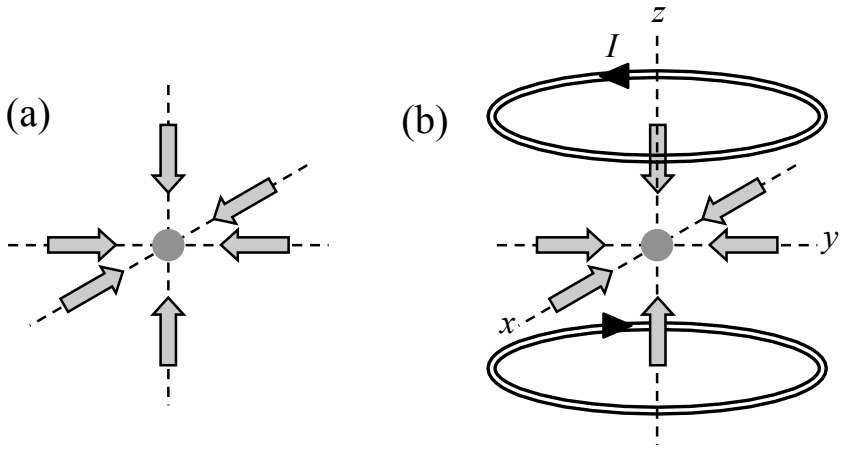

Figure 11.4: (a) Optical molasses. Six laser beams are used to annul the three velocity components of the atom's velocity in both directions. (b) Magneto-optical trap, comprising the optical molasses lasers and a quadrupole magnetic field.

in both directions for all three velocity components (i.e. the $\pm x$, $\pm y$ and $\pm z$ directions), we need a six-beam arrangement as shown in Fig. 11.4(a). This counter-propagating six-beam technique was pioneered by Stephen Chu and co-workers at Bell Laboratories in 1985, and given the name **optical molasses**. "Molasses" is the American word for "treacle", and it gives a good description of how the Doppler cooling force acts like a viscous medium for the trapped atoms.

The optical molasses experiment becomes a magneto-optical trap when magnetic coils are added above and below the intersection point, as shown in Fig. 11.4(b). The current flows in opposite directions through the coils, which produces a **quadrupole field**, where the field at the centre of the apparatus cancels. However, on moving a small distance from the centre, the fields increases according to:

$$B(x, y, z) = B_0(x^2 + y^2 + 4z^2)^{1/2}\,, \tag{11.18}$$

where (x, y, z) is the position relative to the centre. The atoms with $M_J = +1/2$ have a Zeeman energy of $+\mu_B B$, and so experience an increase in energy as they move away from the intersection point of the lasers. In other words, they sit in a potential well, with the minimum at the origin. This has the effect of trapping the atoms close to the origin if their thermal energy is less than the depth of the potential well.[2] The combination of optical molasses and the quadrupole field thus provides a method to cool and trap a gas of atoms at very low temperatures.

11.6 Cooling below the Doppler limit

Careful measurements by Phillips at NIST in 1988 led to the rather startling result that the temperature of the laser-cooled atoms in an optical molasses experiment was substantially less than the Doppler limit given in eqn 11.16. The temperature of the trapped sodium atoms was measured to be 40 μK, that is, six-times smaller than the Doppler limit. Chu and Cohen-Tannoudji soon confirmed this result in independent experiments.

The explanation of the discrepancy comes from realizing that the single-beam mechanism described in Section 11.3 is too simplistic. The counter-propagating laser beams in an optical molasses experiment form an interference pattern, and this leads to a new type of cooling mechanism called **Sisyphus cooling**. The mechanism is named after the character in Greek mythology who was condemned to roll a stone up a hill forever, only for it to roll down again every time he got near the top. This is an analogy for the way Sisyphus cooling works: the atoms repeatedly climb to the top of a potential barrier created by the Stark effect of the interfering laser beams, and then drop to the bottom of the potential barrier after absorption and emission of a photon. The energy loss in the process is taken from the atom's thermal energy.

The detailed mechanism for Sisyphus cooling is too complicated at this level of treatment. The key point is that the minimum temperature that can be achieved is set by the **recoil limit**, rather than

[2] Atoms with $M_J = -1/2$ are not trapped, since their energy decreases as they move away from the centre: it is as if they are on the top of a hill.

		Sodium	Cesium
Laser		rhodamine dye	semiconductor diode
Atomic transition		$3p \rightarrow 3s$	$6p \rightarrow 6s$
Wavelength	λ	589 nm	852 nm
Atomic mass	m	23.0 m_{H}	132.9 m_{H}
Radiative lifetime	τ	16 ns	32 ns
Doppler limit	T_{min}	240 μK	120 μK
Recoil limit	T_{recoil}	2.4 μK	0.2 μK

Table 11.1: Parameters for laser cooling of sodium and cesium atoms. T_{min} and T_{recoil} are the minimum temperature set by the Doppler and photon recoil limits given in eqns 11.16 and 11.20 respectively.

the Doppler limit. The atoms are constantly emitting spontaneous photons of wavelength λ in random directions. The atom recoils each time with momentum h/λ, so it ends up with a random thermal energy given by:

$$\frac{1}{2}k_{\mathrm{B}}T_{\mathrm{recoil}} = \frac{(h/\lambda)^2}{2m} = \frac{h^2}{2m\lambda^2}\,. \tag{11.19}$$

This gives a minimum temperature of:

$$T_{\mathrm{recoil}} = \frac{h^2}{mk_{\mathrm{B}}\lambda^2}\,. \tag{11.20}$$

Table 11.1 compares the key parameters of the sodium and cesium atoms that are frequently used in laser cooling experiments.

In the years since Chu, Cohen-Tannoudji and Phillips' pioneering experiments, the laser cooling techniques have allowed the study of atom-photon interactions with unprecedented precision, and have paved the way for the discovery of Bose-Einstein condensation, as described in the next sections.

11.7 Bose-Einstein condensation

We have seen above how laser cooling techniques can produce a very cold gas of atoms. Despite the extremely low temperatures that are achieved, the motion of the atoms at the focus of the laser beams is still *classical* in terms of statistical mechanics. We now wish to explore what happens when the gas is cooled even further. It turns out that some atoms can undergo a phase transition to a quantum state proposed by Bose and Einstein in 1924–5. In the sections that follow, we first consider the general principles of Bose–Einstein condensation (BEC), and then describe how the experiments to observe BEC in a gas of atoms are carried out.

11.7.1 The concept of Bose–Einstein condensation

The behaviour of a gas of atoms is said to be **classical** if the distribution of energies obeys **Boltzmann statistics**:

$$p(E_i) \propto \exp\left(-\frac{E_i}{k_{\mathrm{B}}T}\right)\,, \tag{11.21}$$

where $p(E_i)$ is probability that the atom is in the quantum state with energy E_i at temperature T. Boltzmann statistics apply at high temperatures when the probability for the occupation of any individual quantum level is small. If we reduce the temperature, the atoms tend to occupy the lowest energy levels of the system. It will therefore eventually be the case that the assumption that the occupancy factor is small no longer applies. In this case, we will have **quantum statistics** rather than classical statistics. It is this regime that we shall be exploring here.

The transition from classical to quantum behaviour occurs at a temperature that is determined by the energy scales of the system. Consider, for example, the specific heat capacity of a gas of diatomic molecules. The variation of the specific heat with temperature is shown schematically in figure 11.5. A diatomic molecule possesses seven degrees of freedom: three translational, two rotational, and two vibrational. As noted in Section 11.2, the classical principle of equipartition of energy states that the thermal energy per molecule per degree of freedom is equal to $\frac{1}{2}k_{\mathrm{B}}T$. Since the heat capacity is equal to $\mathrm{d}E/\mathrm{d}T$, we therefore expect a contribution of $3k_{\mathrm{B}}/2$ for the translational motion, $2k_{\mathrm{B}}/2$ for the rotational

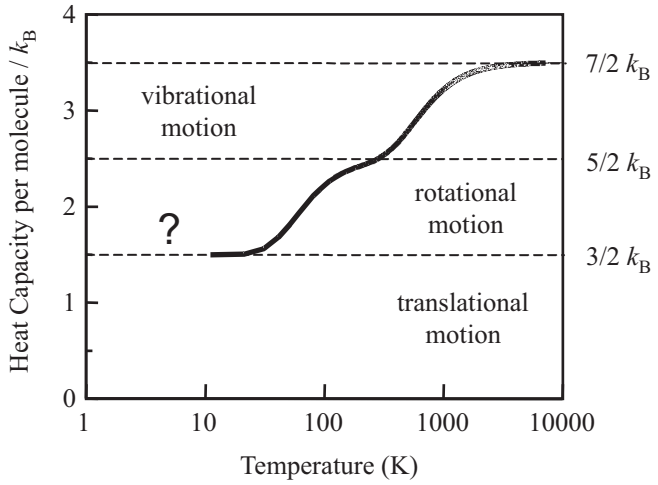

Figure 11.5: Schematic variation of the specific heat capacity of a gas of diatomic molecules with temperature. The rotational and vibrational contributions freeze out at characteristic temperatures, but the freezing out of the translational motion is not normally observed.

motion, and a further $2k_B/2$ for the vibrations, giving $7k_B/2$ in total. This is in fact observed, but only at very high temperatures, as shown in Fig. 11.5.

The reason for the departure of the heat capacity from the classical result is the quantization of the thermal motion. The vibrations of a molecule can be approximated to a simple harmonic oscillator, with quantized energy levels given by:

$$E = (n + 1/2)h\nu_{\rm vib}\,, \tag{11.22}$$

where $\nu_{\rm vib}$ is the vibrational frequency. The classical result will only be obtained if the thermal energy is much greater than the vibrational quanta, that is when

$$k_B T \gg h\nu_{\rm vib}\,. \tag{11.23}$$

With typical values for $\nu_{\rm vib}$ around 10^{13} Hz, the classical behaviour is only observed at temperatures above about 1000 K. At room temperature the vibrational motion is usually "frozen out", as shown in Fig. 11.5. In the same way we expect the rotational motion to freeze out when the thermal energy is comparable to the quantized rotational energy, that is when

$$k_B T \sim \frac{\hbar^2}{I_{\rm rot}}\,, \tag{11.24}$$

where $I_{\rm rot}$ is the moment of inertia about the rotation axis. This typically occurs for $T \sim 50$ K. Thus the rotational motion is usually classical at room temperature, but freezes out at lower temperatures, as indicated in Fig. 11.5.

We are finally left with the translational motion. The **third law of thermodynamics** tells us that the heat capacity must eventually go to zero as we approach absolute zero. However, this is never observed in practice. In any normal gas the attractive forces between the molecules cause liquefaction and solidification long before the quantum effects for the translational motion become important. If, however, we could somehow prevent the gas from condensing, we would eventually expect to observe quantum effects related to the translational motion. This effect was first considered by Einstein in 1924–5, following Bose's work on the statistical mechanics of photons.

A key point in understanding the concept of Bose–Einstein condensation is that we are considering the quantised motion of **non-interacting particles**. The molecules in a gas do *not* normally behave as non-interacting particles: there are attractive forces between them that cause condensation to the liquid or solid phase at low temperatures. These forces can never be turned off, and the only way to make their effect minimal is to keep the molecules far away from each other. This means that the gas density must be very low, which, as we shall see below, makes the temperature required to observe the quantum effect extremely low. This is why is took 70 years to observe Bose–Einstein condensation in a gas.

The phenomenon of Bose–Einstein condensation was described by Einstein in a letter to Paul Ehrenfest in late 1924 as follows:

> Form a certain temperature on, the molecules "condense" without attractive forces, that is, they accumulate at zero velocity. The theory is pretty, but is there some truth to it?[3]

Einstein had to wait 14 years for the beginnings of an answer to his question. The superfluid transition in liquid helium was discovered in 1928 by W.H. Keesom, and in 1938 Fritz London successfully interpreted Keesom's discovery as a Bose–Einstein condensation phenomenon. In the years following London's work, the theory of Bose-Einstein condensation was applied to other condensed matter systems, e.g. superconductors. However, the problem with all of these condensed matter systems is that the particles are *not* "non-interacting". The mere fact that helium is a liquid at the superfluid temperature tells us that there are strong interactions between the atoms over and above any effects due to the quantization of the kinetic energy. For this reason, in 1946 Schrödinger described the modifications to the gas laws caused by the quantum statistics as:[4]

— "satisfactory", because they are negligible at high temperatures and low densities;

— "disappointing", because they occur at such low temperatures and high densities that they are hard to distinguish from other effects;

— "astounding", because the behaviour is completely different to that of a classical system.

In an ideal world we would therefore like to observe the Bose–Einstein condensation in a weakly interacting system (i.e. a gas) so that we can study it in isolation. This was not possible until the new techniques of laser cooling described in the previous sections were developed.

11.7.2 Atomic bosons

Before going into the details of Bose–Einstein condensation, we need to clarify one important point. The quantized behaviour of a gas of identical particles at low temperatures depends on the spin of the particle. Particles with integer spins are called **bosons**, while those with half-integer spins are called **fermions**. Fermions obey the **Pauli exclusion principle**, which says that it is not possible to put more than one particle into a particular quantum state. Bosons, by contrast, do not obey the Pauli principle. There is no limit to the number of particles that can be put into a particular level, which allows the observation of new quantum effects such as BEC.

Atom are composite particles, made up of protons, neutrons, and electrons. These are all spin-1/2 particles, but the composite atom can be either a fermion or a boson depending on its total spin, which can be worked out from:

$$S_{\text{atom}} = S_{\text{electrons}} \oplus I \,, \tag{11.25}$$

where I is the nuclear spin. Since the number of electrons and protons in a neutral atom is equal, it is easy to see that the atom will be a boson if the number of neutrons is an even number, and a fermion if it is odd.

The simplest example to consider is hydrogen. ^{1}H has one proton and one electron, and so we find $S_{\text{atom}} = 0$ or 1. ^{1}H atoms are therefore bosons. Deuterium atoms (^{2}H), by contrast are fermions. Now consider helium. Helium has two common isotopes: ^{4}He and ^{3}He. The ground state of the ^{4}He nucleus is the α-particle with $I = 0$, and the electron ground state also has $S = 0$. (See Chapter 6). Thus the spin of the ^{4}He atom in its ground state is zero, which make it a boson. In ^{3}He atoms, by contrast, the nucleus has two protons and one neutron, with $I = 1/2$ in its ground state The electrons have spin 0 or 1, and so we find $S_{\text{atom}} = 1/2$ or $3/2$, making it a fermion.[5] Note that the number of neutrons is two for ^{4}He and one for ^{3}He, so that our general rule for deciding whether an atom is a boson or a fermion applies.

11.7.3 The condensation temperature

Consider a gas of identical non-interacting bosons of mass m at temperature T. As noted above, the word "non-interacting" is very important here. It implies that the particles are completely free, with only kinetic energy, and no forces between the atoms. In these circumstances the **de Broglie wavelength**

[3] Letter to P. Ehrenfest, 29 November, 1924. An historical discussion of Einstein's work may be found in Pais, A. (1982). *Subtle is the Lord*, Oxford University Press.

[4] See E. Schrödinger, *Statistical Thermodynamics*, Cambridge University Press, 1946.

[5] It is interesting to note that a superfluid phase transition can also be observed for liquid ^{3}He at 2.5 mK, even though the individual atoms are fermions. The ^{3}He atoms pair up to form a bosonic system analogous to the Cooper pairs developed in the BCS (Bardeen–Cooper–Schrieffer) theory of superconductivity. This theory explains how electrons can undergo a superconducting phase transition even though they are fermions.

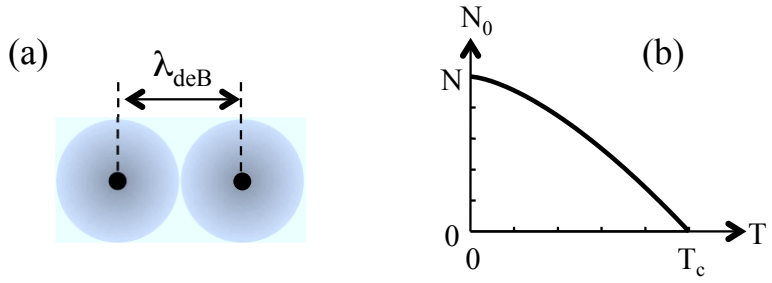

Figure 11.6: (a) Overlapping wave functions of two atoms separated by $\lambda_{\rm deB}$. (b) Number of particles in the Bose-condensed state versus temperature. $T_{\rm c}$ is the condensation temperature given by eqn 11.30.

$\lambda_{\rm deB}$ is determined by the free thermal motion :

$$\frac{p^2}{2m} = \frac{1}{2m}\left(\frac{h}{\lambda_{\rm deB}}\right)^2 = \frac{3}{2}k_{\rm B}T\,. \tag{11.26}$$

This implies that

$$\lambda_{\rm deB} = \frac{h}{\sqrt{3mk_{\rm B}T}}\,. \tag{11.27}$$

The thermal de Broglie wavelength thus increases as T decreases.

The quantum mechanical wave function of a free atom extends over a distance of $\sim \lambda_{\rm deB}$. As $\lambda_{\rm deB}$ increases with decreasing T, a temperature will eventually be reached when the wave functions of neighbouring atoms begin to overlap. This situation is depicted in Fig. 11.6(a). The atoms will interact with each other and coalesce to form a "super atom" with a common wave function. This is the Bose–Einstein condensed state.

The condition for wave function overlap is that the reciprocal of the effective particle volume determined by the de Broglie wavelength should be equal to the particle density. If we have N particles in volume V, this condition can be written:

$$\frac{N}{V} \sim \frac{1}{\lambda_{\rm deB}^3}\,. \tag{11.28}$$

By inserting from eqn 11.27 and solving for T, we find:

$$T_{\rm c} \sim \frac{1}{3}\frac{h^2}{mk_{\rm B}}\left(\frac{N}{V}\right)^{2/3}\,. \tag{11.29}$$

We thus see that the condensation temperature is proportional to $(N/V)^{2/3}$. This shows that low density systems such as gases are expected to have very low transition temperatures, which explains why it has been so difficult to observe BEC in gases until recently.

A rigorous formula for the Bose–condensation temperature $T_{\rm c}$ can be derived by applying the laws of statistical mechanics to the non-interacting boson gas. For a gas of spin-0 bosons, the critical temperature $T_{\rm c}$ is given by:[6]

$$T_{\rm c} = 0.0839\frac{h^2}{mk_{\rm B}}\left(\frac{N}{V}\right)^{2/3}\,. \tag{11.30}$$

Note that this is the same as the intuitive result in eqn 11.29 apart from the numerical factor. As noted previously, the theory of Bose–Einstein condensation was first applied to liquid helium-4. Below $T_{\rm c}$ some of the liquid shows superfluid behaviour, while the remainder remains "normal". On inserting the atom density of ^{4}He into eqn 11.30, we find $T_{\rm c} = 2.7\,{\rm K}$, which is close to the actual superfluid transition temperature of 2.17 K. The discrepancy is a consequence of the fact the ^{4}He atoms in the liquid phase are non truly "non-interacting", and is an example of why Schrödinger described the properties of the quantum gas as "disappointing": the most spectacular effects usually occur in conditions where many other interactions are important.

The picture which emerges from the statistical mechanics of Bose–Einstein condensation is as follows. Above the critical temperature the particles are distributed among the energy states of the system

[6] See, for example, Mandl, *Statistical Physics*, Section 11.6.

according to the **Bose–Einstein distribution**:

$$n_{\mathrm{BE}}(E) = \frac{1}{\exp[(E-\mu)/k_{\mathrm{B}}T] - 1}\,, \tag{11.31}$$

where μ is the chemical potential. In the case that we are considering here, the particles only have kinetic energy with $E = \frac{1}{2}mv^2$, so that the minimum value of E is zero. The chemical potential must therefore be negative to keep n_{BE} well-behaved for all possible values of E. The chemical potential increases with decreasing temperature, and at T_{c} it reaches its maximum value of zero. In these conditions, there is a singularity in eqn 11.31 for the zero-velocity state with $E = 0$, and a phase transition occurs in which a macroscopic fraction of the total number of particles condenses into the ground state. The remainder of the particles continue to be distributed thermally between the finite-velocity states. The fraction of the particles in the zero-velocity state is given by:

$$N_0(T) = N\left[1 - \left(\frac{T}{T_{\mathrm{c}}}\right)^{3/2}\right]\,, \tag{11.32}$$

where N is the total number of particles. This dependence is plotted in Fig. 11.6(b). We see that N_0 is zero at $T = T_{\mathrm{c}}$ and increases to the maximum value of N at $T = 0$.

The description of the system with a macroscopic fraction of the particles in the zero-velocity state and the rest distributed thermally among the finite-velocity states gives rise to the **two-fluid model**. The two fluids correspond to the Bose–Einstein condensed state with $E = 0$, and the "normal" particles with $E > 0$. The total number of particles is written:

$$N = N_{\mathrm{normal}} + N_{\mathrm{condensed}}\,, \tag{11.33}$$

where $N_{\mathrm{condensed}}$ obeys eqn 11.32. This model gives a fairly good description of the behaviour of superfluid liquid ^{4}He and superconductors.

We can relate this behaviour to the discussion of the diatomic gas in Fig. 11.5 in the temperature region indicated by the question mark. Since the number of particles in the zero-velocity state gradually approaches 100% as T goes to zero, the thermal energy of the system goes to zero as $T \to 0$. The heat capacity therefore also goes to zero, and we finally reach consistency with the third law of thermodynamics.

11.8 Experimental techniques for atomic BEC

The conditions required to achieve Bose–Einstein condensation (BEC) in a gas impose severe technical challenges. If we want to observe pure BEC without the complication of other effects such as liquefaction, we have to keep the atoms well apart from each other. This means that the particle density must be small, which in turn implies that the transition temperature is very low.

We have seen in section 11.6 that laser cooling can typically produce temperatures in the range 1–10 μK. This is not quite cold enough. The typical particle density achieved in an optical molasses experiment is around $10^{17}\,\mathrm{m}^{-3}$, which implies condensation temperatures below 100 nK. We therefore have to invent new techniques to observe condensation. The general procedure usually follows three steps:

1. Trap a gas of atoms and cool them towards the recoil-limit temperature using laser-cooling techniques. Compress the gas by increasing the magnetic field.

2. Turn the cooling laser off to permit cooling below the recoil limit.

3. Cool the gas again by evaporative cooling until condensation occurs.

The first step has been discussed previously in section 11.6. The magnetic field has to be ramped up carefully so as not to heat the gas while compressing it. Once the gas has been compressed, the cooling lasers then have to be turned off, since the temperature will not fall below the recoil limit given in eqn 11.20 while the lasers are on.

The final step is called **evaporative cooling**, in analogy to the cooling of a liquid by evaporation. In this technique, the magnetic field strength is gradually turned down in order to reduce the depth of the magnetic potential as shown in Fig. 11.7(b). The fastest-moving atoms now have enough kinetic energy to escape from the trap, leaving the slower ones behind. This causes an overall reduction in the average kinetic energy, which is equivalent to a reduction in the temperature.

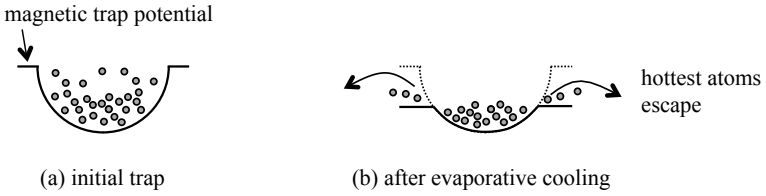

Figure 11.7: Evaporative cooling. (a) The laser-cooled atoms are first compressed in a magnetic trap. (b) The trap potential is then reduced by decreasing the magnetic field strength, so that the hottest atoms can escape. This reduces the temperature, in the same way that evaporation cools a liquid.

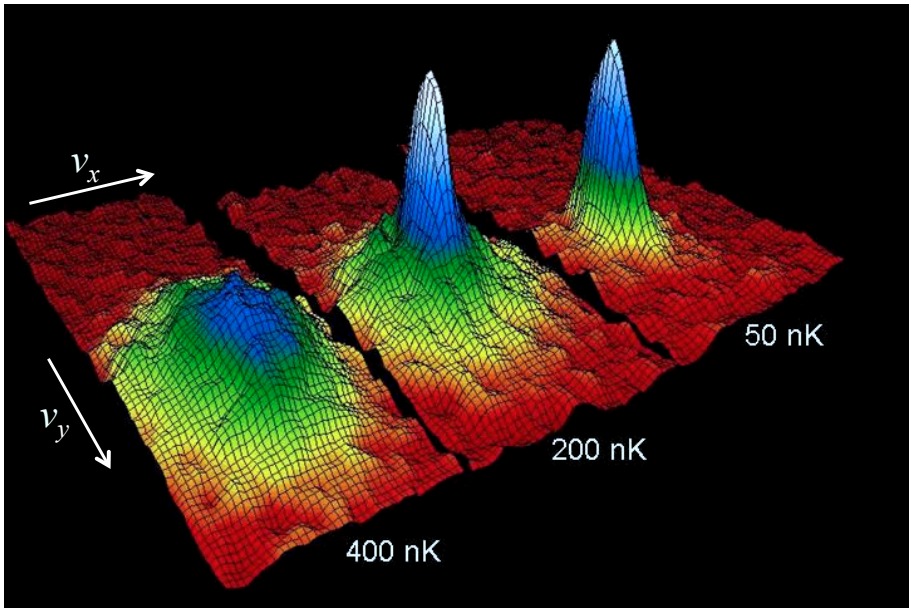

Figure 11.8: Bose-Einstein condensation in rubidium atoms. The three figures show the measured velocity distribution as the gas is cooled through T_c on going from left to right. Above T_c, we have a broad Maxwell–Boltzmann, but as the gas condenses, the fraction of atoms in the zero velocity state at the origin increases dramatically. Source: http://jila.colorado.edu/bec/CornellGroup/index.html.

The first successful observation of Bose–Einstein condensation in an atomic gas was reported by the group of Eric Cornell and Carl Wieman at the JILA Laboratory[7] in the United States in 1995. In their experiments they used ^{87}Rb atoms with a density of about $10^{20}\,\mathrm{m}^{-3}$. This density is eight orders of magnitude smaller than that of liquid helium, and so the condensation temperature calculated from eqn 11.30 is very low: $3.9 \times 10^{-7}\,\mathrm{K}$.[8] The inter-particle distance in the gas is equivalent to about 100 atomic radii. This means that the forces between the atoms are very small, and the BEC effects can be observed in their own right. Similar results were reported by Wolfgang Ketterle and his team at Massachusetts Institute of Technology for a gas of sodium atoms soon afterwards. The ground–breaking nature of these discoveries was recognized by the joint award of the Nobel Prize for Physics in 2001 to Cornell, Ketterle and Wieman.

Bose–Einstein condensation is observed by measuring the velocity distribution of the atoms at the end of the experiment. Figure 11.8 shows some typical data. These pictures are obtained by turning the trapping field off completely and allowing the gas to expand. An image of the gas is taken at a later time, and the velocity distribution can be inferred from the amount of expansion that has occurred. The key point in Fig. 11.8 is that a peak can be seen to appear at the centre as the temperature is lowered. This corresponds to the zero-velocity state, and shows that a macroscopic fraction of the atoms have

[7] Joint Institute for Laboratory Astrophysics, run jointly by the University of Colorado and the National Institute of Standards and Technology (NIST).

[8] The condensation temperature in a magnetic trap differs slightly from the one given in eqn 11.30, because the atoms are subject to the trapping potential. This level of detail need not concern us here.

condensed to the ground state.

In the years since the original observation, BEC has been observed in many other gaseous atomic systems, and this has led to the observation of many other spectacular quantum effects, for example: **atom lasers**. The use of the word "laser" is slightly confusing here, because there is no amplification. It is used to emphasize the difference between the *coherence* of the atomic beam from the condensate and that from a thermal source, in analogy to the difference between the coherence of the light from a laser beam and that from a hot filament. The beam of atoms generated by hot ovens such as the one shown in Fig. 11.3 has a Maxwell–Boltzmann velocity distribution, with random phases between different atoms. The atoms in a beam emanating from a Bose–Einstein condensate, by contrast, are all in phase, because they have a common wave function. This point has been proven by demonstrating that the atomic beams from a condensate can form interference patterns when they overlap.

Further reading

Bransden and Joachain, *Physics of Atoms and Molecules*, sections 15.4–6
Foot, *Atomic physics*, chapters 9 and 10
Fox, *Quantum optics*, chapter 11
Haken, H. and Wolf, *The Physics of Atoms and Quanta*, sections 22.6, 23.11–12.
Mandl, *Statistical Physics*, Section 11.6.

The BEC homepage at the University of Colorado gives interactive tutorial articles on laser cooling and Bose–Einstein condensation. See: http://www.colorado.edu/physics/2000/bec/index.html

Appendix A

The reduced mass

The reduced mass is a very useful concept for dealing with the relative motion of two particles, such as the nucleus and the electron in a hydrogen atom.[1] It allows us to separate the motion of the centre of mass of the whole atom from its internal motion associated with the quantised orbits of the electron around the nucleus. It is the latter that is our concern when solving the Schrödinger equation.

Let $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ be the positive vectors of the two particles, which have masses of m_1 and m_2 respectively. The centre of mass co-ordinate $\boldsymbol{R}$ and the relative co-ordinate $\boldsymbol{r}$ are defined by:

$$\begin{aligned} M\boldsymbol{R} &= m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2\,, \\ \boldsymbol{r} &= \boldsymbol{r}_1 - \boldsymbol{r}_2\,, \end{aligned} \tag{A.1}$$

where $M = (m_1 + m_2)$ is the total mass. As the names suggest, these give the position of the centre of mass and the relative separation of the two particles respectively. The reverse relationships are:

$$\begin{aligned} \boldsymbol{r}_1 &= \boldsymbol{R} + \frac{m_2}{M}\boldsymbol{r}\,, \\ \boldsymbol{r}_2 &= \boldsymbol{R} - \frac{m_1}{M}\boldsymbol{r}\,. \end{aligned} \tag{A.2}$$

We assume that the only force acting on the particles is via their mutual interaction, so that the potential energy $V(r)$ only depends on the separation of the particles. In the case of a hydrogen atom, this is the Coulomb interaction between the two charged particles, with $V(r) = -e^2/4\pi\epsilon_0 r$. In classical mechanics we can write the total energy (i.e. the Hamiltonian) as the sum of the kinetic energies of the particles and the potential energy due to their mutual interaction:

$$H = \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 + V(r)\,. \tag{A.3}$$

It is easily verified from eqn A.2 that:

$$\begin{aligned} (\dot{\boldsymbol{r}}_1)^2 &= \dot{\boldsymbol{R}}^2 + \frac{2m_2}{M}\dot{\boldsymbol{R}}\dot{\boldsymbol{r}} + \left(\frac{m_2}{M}\right)^2 \dot{\boldsymbol{r}}^2 \\ (\dot{\boldsymbol{r}}_2)^2 &= \dot{\boldsymbol{R}}^2 - \frac{2m_1}{M}\dot{\boldsymbol{R}}\dot{\boldsymbol{r}} + \left(\frac{m_1}{M}\right)^2 \dot{\boldsymbol{r}}^2\,. \end{aligned}$$

Hence

$$\begin{aligned} H &= \frac{1}{2}m_1(\dot{\boldsymbol{r}}_1)^2 + \frac{1}{2}m_2(\dot{\boldsymbol{r}}_2)^2 + V(r) \\ &= \frac{1}{2}M\dot{\boldsymbol{R}}^2 + \frac{1}{2}m\dot{\boldsymbol{r}}^2 + V(r)\,, \end{aligned} \tag{A.4}$$

where the **reduced mass** $m = m_1 m_2/M$ is defined by:

$$\frac{1}{m} = \frac{1}{m_1} + \frac{1}{m_2}\,. \tag{A.5}$$

Equation A.4 shows that the energy is equal to the kinetic energy of the centre of mass, plus the energy (i.e. kinetic energy + potential energy) of the relative motion of a particle of mass m, namely the reduced

[1] The reduced mass can also be used in gravitational problems such as planets orbiting the Sun. In that case, the potential energy is $V(r) = -Gm_1m_2/r$.

mass. In other words, we can separate the motion into the free motion of the whole system, plus the internal energy in terms of the relative co-ordinates and the reduced mass.

In quantum mechanics, the Hamiltonian is given by

$$\hat{H} = -\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 + V(r)\,, \tag{A.6}$$

where

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}\,. \tag{A.7}$$

To transform this to the centre of mass and relative co-ordinates, we need to work with the Cartesian co-ordinates:

$$\begin{aligned} x &= x_1 - x_2\,, \\ X &= \frac{m_1}{M}x_1 + \frac{m_2}{M}x_2\,. \end{aligned}$$

We start by finding the first derivatives:

$$\begin{aligned} \frac{\partial}{\partial x_1} &= \frac{\partial X}{\partial x_1}\frac{\partial}{\partial X} + \frac{\partial x}{\partial x_1}\frac{\partial}{\partial x} = \frac{m_1}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial x}\,, \\ \frac{\partial}{\partial x_2} &= \frac{\partial X}{\partial x_2}\frac{\partial}{\partial X} + \frac{\partial x}{\partial x_2}\frac{\partial}{\partial x} = \frac{m_1}{M}\frac{\partial}{\partial X} - \frac{\partial}{\partial x}\,. \end{aligned}$$

This implies that the second derivative with respect to x_1 is:

$$\begin{aligned} \frac{\partial^2}{\partial x_1^2} &= \left(\frac{m_1}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial x}\right)\left(\frac{m_1}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial x}\right)\,, \\ &= \frac{m_1^2}{M^2}\frac{\partial^2}{\partial X^2} + 2\frac{m_1}{M}\frac{\partial^2}{\partial X \partial x} + \frac{\partial^2}{\partial x^2}\,. \end{aligned}$$

Similarly:

$$\frac{\partial^2}{\partial x_2^2} = \frac{m_2^2}{M^2}\frac{\partial^2}{\partial X^2} - 2\frac{m_2}{M}\frac{\partial^2}{\partial X \partial x} + \frac{\partial^2}{\partial x^2}\,.$$

Therefore:

$$-\frac{\hbar^2}{2m_1}\frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m_2}\frac{\partial^2}{\partial x_2^2} = -\frac{\hbar^2}{2M}\frac{\partial^2}{\partial X^2} - \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\,,$$

where m is the reduced mass defined in eqn A.5. Similar results can be derived for the y and z components, leading to:

$$\hat{H} = \hat{H}_R + \hat{H}_r\,, \tag{A.8}$$

where

$$\begin{aligned} \hat{H}_R &= -\frac{\hbar^2}{2M}\nabla_R^2\,, \\ \hat{H}_r &= -\frac{\hbar^2}{2m}\nabla_r^2 + V(r)\,. \end{aligned} \tag{A.9}$$

This shows that the Hamiltonian is the sum of:

- The Hamiltonian $\hat{H}_R$ of a free particle of mass M with position co-ordinates of the centre of mass;
- The Hamiltonian $\hat{H}_r$ that describes the relative motion of the two particles, behaving as if they had mass m, namely the reduced mass.

This is our final result. It shows that we can separate the motion of hydrogenic atoms into the motion of the centre of mass, that moves freely throughout space, and the internal motion, that is governed by the potential energy $V(r)$ and acts like a particle of mass m. Hence the mass m that appears in the Bohr model in Section 2.1 and in the hydrogen Schrödinger equation in Section 2.2 is the reduced mass. This separation works for any central potential that depends only on the particle separation r.

Appendix B

Mathematical solutions of the radial equation

This appendix deals with the more mathematical aspects of the Schrödinger equation for hydrogen that were omitted from the main discussion in Chapter 2.

B.1 The angular equation

The eigenfunctions of the angular momentum operator are found by solving equation 2.32, namely:

$$\hat{\boldsymbol{L}}^2 F(\theta,\phi) \equiv -\hbar^2 \left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right] F(\theta,\phi) = CF(\theta,\phi)\,. \tag{B.1}$$

For reasons that will become clearer later, the constant C is usually written in the form:

$$C = l(l+1)\hbar^2\,. \tag{B.2}$$

At this stage, l can take any value, real or complex. We can separate the variables by writing:

$$F(\theta,\phi) = \Theta(\theta)\Phi(\phi)\,. \tag{B.3}$$

On substitution into eqn 2.32 and cancelling the common factor of $\hbar^2$, we find:

$$-\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)\Phi - \frac{1}{\sin^2\theta}\Theta\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2} = l(l+1)\Theta\Phi\,. \tag{B.4}$$

Multiply by $-\sin^2\theta/\Theta\Phi$ and re-arrange to obtain:

$$\frac{\sin\theta}{\Theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right) + \sin^2\theta\, l(l+1) = -\frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2}\,. \tag{B.5}$$

The left hand side is a function of θ only, while the right hand side is a function of ϕ only. The equation must hold for all values of the θ and ϕ and hence both sides must be equal to a constant. On writing this arbitrary separation constant m^2, we then find:

$$\sin\theta\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right) + l(l+1)\sin^2\theta\,\Theta = m^2\Theta\,, \tag{B.6}$$

and

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2} = -m^2\Phi\,. \tag{B.7}$$

The equation in ϕ is easily solved to obtain:

$$\Phi(\phi) = A\mathrm{e}^{\mathrm{i}m\phi}\,. \tag{B.8}$$

The wave function must have a single value for each value of ϕ, and hence we require:

$$\Phi(\phi + 2\pi) = \Phi(\phi)\,, \tag{B.9}$$

which requires that the separation constant m must be an integer. Using this fact in eqn B.6, we then have to solve

$$\sin\theta\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right)+[l(l+1)\sin^2\theta-m^2]\,\Theta=0\,, \tag{B.10}$$

with the constraint that m must be an integer. On making the substitution $u=\cos\theta$ and writing $\Theta(\theta)=P(u)$, eqn B.10 becomes:

$$\frac{\mathrm{d}}{\mathrm{d}u}\left((1-u^2)\frac{\mathrm{d}P}{\mathrm{d}u}\right)+\left[l(l+1)-\frac{m^2}{1-u^2}\right]P=0\,. \tag{B.11}$$

Equation B.11 is known as either the Legendre equation or the associated Legendre equation, depending on whether m is zero or not. Solutions only exist if l is an integer $\geq|m|$ and $P(u)$ is a polynomial function of u. This means that the solutions to eqn B.10 are of the form:

$$\Theta(\theta)=P_l^m(\cos\theta)\,, \tag{B.12}$$

where $P_l^m(\cos\theta)$ is a polynomial function in $\cos\theta$ called the (associated) **Legendre polynomial** function.

Putting this all together, we then find:

$$F(\theta,\phi)=\text{normalization constant}\times P_l^m(\cos\theta)\,\mathrm{e}^{\mathrm{i}m\phi}\,, \tag{B.13}$$

where m and l are integers, and m can have values from $-l$ to $+l$. The correctly normalized functions are called the **spherical harmonic** functions $Y_{lm}(\theta,\phi)$.

It is apparent from eqns B.1 and B.2 that the spherical harmonics satisfy:

$$\hat{\boldsymbol{L}}^2Y_{lm}(\theta,\phi)=l(l+1)\hbar^2Y_{lm}(\theta,\phi)\,. \tag{B.14}$$

Furthermore, on substituting from eqn 2.33, it is also apparent that

$$\hat{L}_zY_{lm}(\theta,\phi)=m\hbar Y_{lm}(\theta,\phi)\,. \tag{B.15}$$

The integers l and m that appear here are called the orbital and magnetic quantum numbers respectively. Some of the spherical harmonic functions are listed in Table 2.2. Equations B.14–B.15 show that the magnitude of the angular momentum and its z-component are equal to $\sqrt{l(l+1)}\hbar$ and $m\hbar$ respectively, as consistent with Fig. 2.3.

B.2 The radial equation

The radial wave equation for hydrogen is given from eqn 2.30 as:

$$-\frac{\hbar^2}{2m}\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}R(r)}{\mathrm{d}r}\right)+\frac{\hbar^2l(l+1)}{2mr^2}R(r)-\frac{Ze^2}{4\pi\epsilon_0r}R(r)=ER(r)\,, \tag{B.16}$$

where l is an integer ≥ 0. We first put this in a more user-friendly form by introducing the dimensionless radius ρ according to:

$$\rho=\left(\frac{8m|E|}{\hbar^2}\right)^{1/2}r\,. \tag{B.17}$$

The modulus sign around E is important here because we are seeking bound solutions where E is negative. The radial equation now becomes:

$$\frac{\mathrm{d}^2R}{\mathrm{d}\rho^2}+\frac{2}{\rho}\frac{\mathrm{d}R}{\mathrm{d}\rho}+\left(\frac{\lambda}{\rho}-\frac{1}{4}-\frac{l(l+1)}{\rho^2}\right)R=0\,, \tag{B.18}$$

where

$$\lambda=\frac{1}{4\pi\epsilon_0}\frac{Ze^2}{\hbar}\left(\frac{m}{2|E|}\right)^{1/2}\,. \tag{B.19}$$

We first consider the behaviour at $\rho\to\infty$, where eqn B.18 reduces to:

$$\frac{\mathrm{d}^2R}{\mathrm{d}\rho^2}-\frac{1}{4}R=0\,. \tag{B.20}$$

This has solutions of $e^{\pm\rho/2}$. The $e^{+\rho/2}$ solution cannot be normalized and is thus excluded, which implies that $R(\rho) \sim e^{-\rho/2}$.

Now consider the behaviour for $\rho \to 0$, where the dominant terms in eqn B.18 are:

$$\frac{d^2R}{d\rho^2} + \frac{2}{\rho}\frac{dR}{d\rho} - \frac{l(l+1)}{\rho^2}R = 0\,, \tag{B.21}$$

with solutions $R(\rho) = \rho^l$ or $R(\rho) = \rho^{-(l+1)}$. The latter diverges at the origin and is thus unacceptable.

The consideration of the asymptotic behaviours suggests that we should look for general solutions of the radial equation with $R(\rho)$ in the form:

$$R(\rho) = L(\rho)\,\rho^l\,e^{-\rho/2}\,. \tag{B.22}$$

On substituting into eqn B.18 we find:

$$\frac{d^2L}{d\rho^2} + \left(\frac{2l+2}{\rho} - 1\right)\frac{dL}{d\rho} + \frac{\lambda - l - 1}{\rho}\,L = 0\,. \tag{B.23}$$

We now look for a series solution of the form:

$$L(\rho) = \sum_{k=0}^{\infty} a_k \rho^k\,. \tag{B.24}$$

Substitution into eqn B.23 yields:

$$\sum_{k=0}^{\infty}\left[k(k-1)a_k\rho^{k-2} + \left(\frac{2l+2}{\rho} - 1\right)ka_k\rho^{k-1} + \frac{\lambda - l - 1}{\rho}a_k\rho^k\right] = 0\,, \tag{B.25}$$

which can be re-written:

$$\sum_{k=0}^{\infty}\left[(k(k-1) + 2k(l+1))a_k\rho^{k-2} + (\lambda - l - 1 - k)a_k\rho^{k-1}\right] = 0\,, \tag{B.26}$$

or alternatively:

$$\sum_{k=0}^{\infty}\left[((k+1)k + 2(k+1)(l+1))a_{k+1}\rho^{k-1} + (\lambda - l - 1 - k)a_k\rho^{k-1}\right] = 0\,. \tag{B.27}$$

This will be satisfied if

$$((k+1)k + 2(k+1)(l+1))a_{k+1} + (\lambda - l - 1 - k)a_k = 0\,, \tag{B.28}$$

which implies:

$$\frac{a_{k+1}}{a_k} = \frac{-\lambda + l + 1 + k}{(k+1)(k+2l+2)}\,. \tag{B.29}$$

At large k we have:

$$\frac{a_{k+1}}{a_k} \sim \frac{1}{k}\,. \tag{B.30}$$

Now the series expansion of e^ρ is

$$e^\rho = 1 + \rho + \frac{\rho^2}{2!} + \cdots \frac{\rho^k}{k!} + \cdots\,, \tag{B.31}$$

which has the same limit for a_{k+1}/a_k. With $R(\rho)$ given by eqn B.22, we would then have a dependence of $e^{+\rho}\cdot e^{-\rho/2} = e^{+\rho/2}$, which is unacceptable. We therefore conclude that the series expansion must terminate for some value of k. Let n_r be the value of k for which the series terminates. It then follows that $a_{n_r+1} = 0$, which implies:

$$-\lambda + l + 1 + n_r = 0\,, \qquad n_r \geq 0\,, \tag{B.32}$$

or

$$\lambda = l + 1 + n_r\,. \tag{B.33}$$

We now introduce the **principal quantum number** n according to:

$$n = n_r + l + 1\,. \tag{B.34}$$

It follows that:

1. n is an integer,
2. $n \geq l + 1$,
3. $\lambda = n$.

The first two points establish the general rules for the quantum numbers n and l. The third one fixes the energy. On inserting $\lambda = n$ into eqn B.19 and remembering that E is negative, we find:

$$E_n = -\frac{me^4}{(4\pi\epsilon_0)^2 2\hbar^2}\,\frac{Z^2}{n^2}\,. \tag{B.35}$$

This is the usual Bohr result. The wave functions are of the form given in eqn B.22:

$$R(\rho) = \rho^l\, L(\rho)\, \mathrm{e}^{-\rho/2}\,. \tag{B.36}$$

The polynomial series $L(\rho)$ that satisfies eqn B.23 is known as an **associated Laguerre function**. On substituting for ρ from eqn B.17 with $|E|$ given by eqn B.35, we then obtain:

$$R(r) = \text{normalization constant} \times \text{Laguerre polynomial in } r \times r^l \mathrm{e}^{-r/a} \tag{B.37}$$

as before (cf. eqn 2.43), with

$$a = \left(\frac{\hbar^2}{2m|E|}\right)^{1/2} = \frac{4\pi\epsilon_0\hbar^2}{me^2}\,\frac{n}{Z} \equiv \frac{n}{Z} a_{\mathrm{H}}\,, \tag{B.38}$$

where a_{H} is the Bohr radius of hydrogen.

Appendix C

Helium energy integrals

The concept of exchange integrals was introduced in § 6.4 in the discussion of the energy levels of helium in Chapter 6. Our task here is to evaluate the three terms that appear in the gross structure energy E:

$$E = E_1 + E_2 + E_{12}\,, \tag{C.1}$$

where the energies are defined in eqns 6.13 and 6.14.

We restrict ourselves to configurations of the type (1s,nl), since these are the ones that give rise to the excited states that are observed in the optical spectra. From eqn 6.4 we see that spatial part of the wave function is given by:

$$\Psi(\boldsymbol{r}_1,\boldsymbol{r}_2) = \frac{1}{\sqrt{2}}\Big(u_{1s}(\boldsymbol{r}_1)u_{nl}(\boldsymbol{r}_2) \pm u_{nl}(\boldsymbol{r}_1)u_{1s}(\boldsymbol{r}_2)\Big)$$

where we take the + sign for singlets with $S = 0$ and the − sign for triplets with $S = 1$.

We first tackle E_1, with $\hat{H}_1$ defined in eqn 6.10:

$$\begin{aligned} E_1 &= \iint \Psi^*\,\hat{H}_1\,\Psi\,\mathrm{d}^3\boldsymbol{r}_1\mathrm{d}^3\boldsymbol{r}_2 \\ &= \frac{1}{2}\iint \Big(u^*_{1s}(\boldsymbol{r}_1)\,u^*_{nl}(\boldsymbol{r}_2) \pm u^*_{nl}(\boldsymbol{r}_1)\,u^*_{1s}(\boldsymbol{r}_2)\Big) \\ &\qquad \hat{H}_1\Big(u^*_{1s}(\boldsymbol{r}_1)\,u^*_{nl}(\boldsymbol{r}_2) \pm u^*_{nl}(\boldsymbol{r}_1)\,u^*_{1s}(\boldsymbol{r}_2)\Big)\,\mathrm{d}^3\boldsymbol{r}_1\,\mathrm{d}^3\boldsymbol{r}_2\,, \end{aligned}$$

where the + sign applies for singlet states and the − sign for triplets. This splits into four integrals:

$$\begin{aligned} E_1 &= \frac{1}{2}\iint u^*_{1s}(\boldsymbol{r}_1)u^*_{nl}(\boldsymbol{r}_2)\hat{H}_1u_{1s}(\boldsymbol{r}_1)u_{nl}(\boldsymbol{r}_2)\mathrm{d}^3\boldsymbol{r}_1\mathrm{d}^3\boldsymbol{r}_2 \\ &\quad +\frac{1}{2}\iint u^*_{nl}(\boldsymbol{r}_1)u^*_{1s}(\boldsymbol{r}_2)\hat{H}_1u_{nl}(\boldsymbol{r}_1)u_{1s}(\boldsymbol{r}_2)\mathrm{d}^3\boldsymbol{r}_1\mathrm{d}^3\boldsymbol{r}_2 \\ &\quad \pm\frac{1}{2}\iint u^*_{1s}(\boldsymbol{r}_1)u^*_{nl}(\boldsymbol{r}_2)\hat{H}_1u_{nl}(\boldsymbol{r}_1)u_{1s}(\boldsymbol{r}_2)\,\mathrm{d}^3\boldsymbol{r}_1\mathrm{d}^3\boldsymbol{r}_2 \\ &\quad \pm\frac{1}{2}\iint u^*_{nl}(\boldsymbol{r}_1)u^*_{1s}(\boldsymbol{r}_2)\hat{H}_1u_{1s}(\boldsymbol{r}_1)u_{nl}(\boldsymbol{r}_2)\mathrm{d}^3\boldsymbol{r}_1\mathrm{d}^3\boldsymbol{r}_2\,. \end{aligned}$$

We now use the fact that $u_{nl}(\boldsymbol{r}_1)$ is an eigenstate of $\hat{H}_1$:

$$\hat{H}_1\,u_{nl}(\boldsymbol{r}_1) = E_{nl}\,u_{nl}(\boldsymbol{r}_1)\,,$$

and that $\hat{H}_1$ has no effect on $\boldsymbol{r}_2$, to obtain:

$$\begin{aligned} E_1 &= \frac{1}{2}E_{1s}\int u^*_{1s}(\boldsymbol{r}_1)u_{1s}(\boldsymbol{r}_1)\mathrm{d}^3\boldsymbol{r}_1\int u^*_{nl}(\boldsymbol{r}_2)u_{nl}(\boldsymbol{r}_2)\mathrm{d}^3\boldsymbol{r}_2 \\ &\quad +\frac{1}{2}E_{nl}\int u^*_{nl}(\boldsymbol{r}_1)u_{nl}(\boldsymbol{r}_1)\mathrm{d}^3\boldsymbol{r}_1\int u^*_{1s}(\boldsymbol{r}_2)u_{1s}(\boldsymbol{r}_2)\mathrm{d}^3\boldsymbol{r}_2 \\ &\quad \pm\frac{1}{2}E_{nl}\int u^*_{1s}(\boldsymbol{r}_1)u_{nl}(\boldsymbol{r}_1)\,\mathrm{d}^3\boldsymbol{r}_1\int u^*_{nl}(\boldsymbol{r}_2)u_{1s}(\boldsymbol{r}_2)\mathrm{d}^3\boldsymbol{r}_2 \\ &\quad \pm\frac{1}{2}E_{1s}\int u^*_{nl}(\boldsymbol{r}_1)u_{1s}(\boldsymbol{r}_1)\mathrm{d}^3\boldsymbol{r}_1\int u^*_{1s}(\boldsymbol{r}_2)u_{nl}(\boldsymbol{r}_2)\mathrm{d}^3\boldsymbol{r}_2 \\ &= \frac{1}{2}E_{1s} + \frac{1}{2}E_{nl} + 0 + 0\,. \end{aligned}$$

The integrals in the first two terms are unity because the u_{nl} wave functions are normalized, while the last two terms are zero by orthogonality.

The evaluation of E_2 follows a similar procedure:

$$\begin{aligned}
E_2 &= \iint \Psi^* \hat{H}_2 \Psi \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \,, \\
&= +\frac{1}{2}\iint u^*_{1s}(\boldsymbol{r}_1) u^*_{nl}(\boldsymbol{r}_2) \hat{H}_2 u_{1s}(\boldsymbol{r}_1) u_{nl}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&\quad +\frac{1}{2}\iint u^*_{nl}(\boldsymbol{r}_1) u^*_{1s}(\boldsymbol{r}_2) \hat{H}_2 u_{nl}(\boldsymbol{r}_1) u_{1s}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&\quad \pm\frac{1}{2}\iint u^*_{1s}(\boldsymbol{r}_1) u^*_{nl}(\boldsymbol{r}_2) \hat{H}_2 u_{nl}(\boldsymbol{r}_1) u_{1s}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&\quad \pm\frac{1}{2}\iint u^*_{nl}(\boldsymbol{r}_1) u^*_{1s}(\boldsymbol{r}_2) \hat{H}_2 u_{nl}(\boldsymbol{r}_1) u_{1s}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&= +\frac{1}{2}E_{nl} + \frac{1}{2}E_{1s} + 0 + 0 \,.
\end{aligned}$$

Finally, we have to evaluate the Coulomb repulsion term, with $\hat{H}_{12}$ defined in eqn 6.11:

$$\begin{aligned}
E_{12} &= \iint \Psi^* \, \hat{H}_{12} \, \Psi \, \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&= \iint \Psi^* \, \frac{e^2}{4\pi\epsilon_0 r_{12}} \, \Psi \, \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&= \frac{1}{2}\iint \Big(u^*_{1s}(\boldsymbol{r}_1)\, u^*_{nl}(\boldsymbol{r}_2) \pm u^*_{nl}(\boldsymbol{r}_1)\, u^*_{1s}(\boldsymbol{r}_2) \Big) \frac{e^2}{4\pi\epsilon_0 r_{12}} \\
&\qquad \Big(u^*_{1s}(\boldsymbol{r}_1)\, u^*_{nl}(\boldsymbol{r}_2) \pm u^*_{nl}(\boldsymbol{r}_1)\, u^*_{1s}(\boldsymbol{r}_2) \Big) \, \mathrm{d}^3\boldsymbol{r}_1 \, \mathrm{d}^3\boldsymbol{r}_2 \,,
\end{aligned}$$

where again the + sign applies for singlet states and the − sign for triplets. The four terms are:

$$\begin{aligned}
E_{12} &= +\frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\iint u^*_{1s}(\boldsymbol{r}_1) u^*_{nl}(\boldsymbol{r}_2) \frac{1}{r_{12}} u_{1s}(\boldsymbol{r}_1) u_{nl}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&\quad +\frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\iint u^*_{nl}(\boldsymbol{r}_1) u^*_{1s}(\boldsymbol{r}_2) \frac{1}{r_{12}} u_{nl}(\boldsymbol{r}_1) u_{1s}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&\quad \pm\frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\iint u^*_{1s}(\boldsymbol{r}_1) u^*_{nl}(\boldsymbol{r}_2) \frac{1}{r_{12}} u_{nl}(\boldsymbol{r}_1) u_{1s}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&\quad \pm\frac{1}{2}\frac{e^2}{4\pi\epsilon_0}\iint u^*_{nl}(\boldsymbol{r}_1) u^*_{1s}(\boldsymbol{r}_2) \frac{1}{r_{12}} u_{1s}(\boldsymbol{r}_1) u_{nl}(\boldsymbol{r}_2) \mathrm{d}^3\boldsymbol{r}_1 \mathrm{d}^3\boldsymbol{r}_2 \\
&= +\frac{D}{2} + \frac{D}{2} \pm \frac{J}{2} \pm \frac{J}{2} \,,
\end{aligned}$$

where D and J are given by eqns 6.18 and 6.19 respectively.

The total energy is thus given by

$$\begin{aligned}
E &= E_{1s} + E_{nl} + D \pm J \\
&= -4R_{\mathrm{H}} - 4R_{\mathrm{H}}/n^2 + D \pm J \,,
\end{aligned}$$

where the + sign applies to singlets and the − sign to triplets. (cf eqn 6.20 with $n_1 = 1$ and $n_2 = n$.)

Appendix D

Perturbation theory of the Stark effect

This appendix gives an explanation of the quadratic and linear Stark shifts by perturbation theory. The basic phenomena were described in § 8.4.1 of Chapter 8. We focus specifically on the quadratic shift in an alkali atom, and the linear shift in hydrogen.

D.1 Quadratic Stark shifts

The energy shift caused by the quadratic Stark effect can be evaluated by applying perturbation theory. The perturbation to the energy of the electrons by a field $\boldsymbol{\mathcal{E}}$is of the form:

$$\begin{aligned} H' &= -\sum_i (-e\boldsymbol{r}_i)\cdot\boldsymbol{\mathcal{E}}\,, \\ &= e\mathcal{E}\sum_i z_i\,, \end{aligned} \tag{D.1}$$

where the field is assumed to point in the $+z$ direction. This is just the sum of the interaction energies of the electron dipoles with he electric field. In principle, the sum is over all the electrons, but in practice, we need only consider the valence electrons, because the electrons in closed shells are very strongly bound to the nucleus and are therefore very hard to perturb. In writing eqn D.1, we take, as always, $\boldsymbol{r}_i$ to be the relative displacement of the electron with respect to the nucleus.

For simplicity, we shall just consider the case of alkali atoms which possess only one valence electron. In this case, the perturbation to the valence electron caused by the field reduces to:

$$H' = e\mathcal{E}z\,. \tag{D.2}$$

The *first-order* energy shift is given by:

$$\Delta E = \langle\psi|H'|\psi\rangle = e\mathcal{E}\langle\psi|z|\psi\rangle\,, \tag{D.3}$$

where

$$\langle\psi|z|\psi\rangle = \iiint_{\text{all space}} \psi^* z\,\psi\,\mathrm{d}^3\boldsymbol{r}\,. \tag{D.4}$$

Now unperturbed atomic states have definite parities. (See discussion in Section 3.4.) The product $\psi^*\psi = |\psi^2|$ is therefore an even function, while z is an odd function. It is therefore apparent that

$$\langle\psi|z|\psi\rangle = \iiint_{\text{all space}} (\text{even function})\ \times\ (\text{odd function})\,\mathrm{d}^3\boldsymbol{r} = 0\,.$$

The first-order energy shift is therefore zero, which explain why the energy shift is quadratic in the field, rather than linear.

The quadratic energy shift can be calculated by *second-order* perturbation theory. In general, the energy shift of the ith state predicted by second-order perturbation theory is given by:

$$\Delta E_i = \sum_{j\neq i} \frac{|\langle\psi_i|H'|\psi_j\rangle|^2}{E_i - E_j}\,, \tag{D.5}$$

where the summation runs over all the other states of the system, and E_i and E_j are the unperturbed energies of the states. The condition of validity is that the magnitude of the perturbation, namely $|\langle\psi_i|H'|\psi_j\rangle|$, should be small compared to the unperturbed energy splittings. For the Stark shift of the valence electron of an alkali atom, this becomes:

$$\Delta E_i = e^2\mathcal{E}^2 \sum_{j\neq i} \frac{|\langle\psi_i|z|\psi_j\rangle|^2}{E_i - E_j} \,. \tag{D.6}$$

We see immediately that the shift is expected to quadratic in the field, which is indeed the case for most atoms.

As a specific example, we consider sodium, which has a single valence electron in the 3s shell. We first consider the ground state $3s\,{}^2S_{1/2}$ term. The summation in eqn D.6 runs over all the excited states of sodium, namely the 3p, 3d, 4s, 4p, ... states. Now in order that the matrix element $\langle\psi_i|z|\psi_j\rangle$ should be non-zero, it is apparent that the states i and j must *opposite* parities. In this case, we would have:

$$\langle\psi_i|z|\psi_j\rangle = \iiint_{\text{all space}} (\text{even/odd parity}) \times (\text{odd parity}) \times (\text{odd/even parity})\, \mathrm{d}^3\boldsymbol{r} \neq 0 \,,$$

since the integrand is an even function. On the other hand, if the states have the same parities, we have:

$$\langle\psi_i|z|\psi_j\rangle = \iiint_{\text{all space}} (\text{even/odd parity}) \times (\text{odd parity}) \times (\text{even/odd parity})\, \mathrm{d}^3\boldsymbol{r} = 0 \,,$$

since the integrand is an odd function. Since the parity varies as $(-1)^l$, the s and d states do not contribute to the Stark shift of the 3s state, and the summation in eqn D.6 is only over the p and f excited states. Owing to the energy difference factor in the denominator, the largest perturbation to the 3s state will arise from the first excited state, namely the 3p state. Since this lies above the 3s state, the energy difference in the denominator is negative, and the energy shift is therefore negative. Indeed, it is apparent that the quadratic Stark shift of the ground state of an atom will always be negative, since the denominator will be negative for all the available states of the system. This implies that the Stark effect will always correspond to a *red shift* for the ground state level.

There is no easy way to calculate the size of the energy shift, but we can give a rough order of magnitude estimate. If we neglect the contributions of the even parity excited states above the 3p state, the energy shift will be given by:

$$\Delta E_{3s} \approx -e^2\mathcal{E}^2 \frac{|\langle\psi_{3s}|z|\psi_{3p}\rangle|^2}{E_{3p} - E_{3s}} \,.$$

The expectation value of z over the atom must be smaller than a, where a is the atomic radius of sodium, namely 0.18 nm. Hence with $E_{3p} - E_{3s} = 2.1\,\text{eV}$, we then have:

$$\Delta E_{3s} \lesssim -\frac{e^2a^2}{E_{3p} - E_{3s}}\mathcal{E}^2 \,.$$

On introducing the atomic polarizability defined in eqn 8.29, we then find that $\alpha_{3s} \lesssim 3.2\times10^{-20}\,\text{eV m}^2\,\text{V}^{-2}$. This predicts a shift of $\lesssim -1\times10^{-5}\,\text{eV}$ ($-0.08\,\text{cm}^{-1}$) in a field of $2.5\times10^7\,\text{V/m}$, which compares reasonably well with the experimental value of $-0.6\times10^{-5}\,\text{eV}$ ($-0.05\,\text{cm}^{-1}$).

The order of magnitude calculation given above can also provide a useful estimation of the field strength at which the second-order perturbation approximation breaks down. This will occur when the magnitude of the perturbation become comparable to the unperturbed energy splitting, that is when:

$$e\mathcal{E}|\langle\psi_{3s}|z|\psi_{3p}\rangle| \sim (E_{3p} - E_{3s}) \,.$$

On setting $|\langle\psi_{3s}|z|\psi_{3p}\rangle| = a$ as before, we find $\mathcal{E} \sim 10^{10}\,\text{V/m}$, which is an extremely large field. The second-order perturbation approach will therefore be a good approximation in most practical situations.

Now consider the Stark shift of the 3p state. The 3p state has odd parity, and so the non-zero contributions in eqn D.6 will now arise from the even parity ns and nd states:

$$\Delta E_{3p} = e^2\mathcal{E}^2 \left(\frac{|\langle\psi_{3p}|z|\psi_{3s}\rangle|^2}{E_{3p} - E_{3s}} + \frac{|\langle\psi_{3p}|z|\psi_{3d}\rangle|^2}{E_{3p} - E_{3d}} + \frac{|\langle\psi_{3p}|z|\psi_{4s}\rangle|^2}{E_{3p} - E_{4s}} + \cdots \right) .$$

The first term gives a positive shift, while all subsequent terms are negative. Therefore, it is not immediately obvious that the Stark shift of excited states like the 3p state will be negative. However, since

the energy difference of the excited states tends to get smaller as we go up the ladder of levels, it will generally be the case that the negative terms dominate, and we have a red shift as for the ground state. Moreover, the red shift is generally expected to be larger than that of the ground state for the same reason (i.e. the smaller denominator). In the case of the 3p state of sodium, the largest contribution comes from the 3d state which lies 1.51 eV above the 3p state, even though the 4s state is closer (relative energy +1.09 eV). This is because of the smaller value of the matrix element for the s states.

D.2 Linear Stark effect

The second-order energy shift given by eqn D.6 diverges if an atom possesses degenerate states with opposite parities. This is the case for the l states of hydrogen with the same n. A new approach to calculate the Stark shift must then be taken based on *degenerate* perturbation theory.

Consider first the 1s ground state of hydrogen. This level is unique, and hence the second-order perturbation approach is valid. A small quadratic red-shift therefore occurs, as discussed in the previous sub-section.

Now consider the $n = 2$ shell, which has four levels, namely the $m = 0$ level from the 2s term, and the $m = -1$, 0, and +1 levels of the 2p term. In the absence of an applied field, these four levels are degenerate. If the atom is in the $n = 2$ shell, it is equally likely to be in any of the four degenerate levels. We must therefore write its wave function as:

$$\psi_{n=2} = \sum_{i=1}^{4} c_i \psi_i \,, \tag{D.7}$$

where the subscript i identifies the quantum numbers $\{n, l, m\}$, that is:

$$\psi_1 \equiv \psi_{2,0,0} \,; \qquad \psi_2 \equiv \psi_{2,1,-1} \,; \qquad \psi_3 \equiv \psi_{2,1,0} \,; \qquad \psi_4 \equiv \psi_{2,1,+1} \,.$$

The first-order energy shift from eqn D.3 becomes:

$$\Delta E = e\mathcal{E} \sum_{i,j} c_i c_j \langle \psi_i | z | \psi_j \rangle \,. \tag{D.8}$$

Unlike the case of the ground state, we can see from parity arguments that some of the matrix elements are non-zero. For example, ψ_1 has even parity, but ψ_3 has odd parity. We therefore have:

$$\begin{aligned} \langle \psi_1 | z | \psi_3 \rangle &= \iiint_{\text{all space}} \psi_1^* \, z \, \psi_3 \, \mathrm{d}^3\boldsymbol{r} \,, \\ &= \iiint_{\text{all space}} (\text{even parity}) \times (\text{odd parity}) \times (\text{odd parity}) \, \mathrm{d}^3\boldsymbol{r} \,, \\ &\neq 0 \,. \end{aligned}$$

This implies that we can observe a *linear* shift of the levels with the field. It turns out that $\langle \psi_1 | z | \psi_3 \rangle$ is the only non-zero matrix element. This is because the perturbation $H' = e\mathcal{E}z$ commutes with $\hat{L}_z$, and so the only non-zero matrix elements are those between states with the same m value but opposite parity, that is, between the two $m = 0$ levels derived from the 2s and 2p states.

It can easily be evaluated from the hydrogenic wave functions of the $n = 2$ levels given in Tables 2.2 and 2.3 that:

$$\langle \psi_1 | z | \psi_3 \rangle = -3a_0 \,,$$

where a_0 is the Bohr radius of hydrogen. We then deduce that the field splits the $n = 2$ shell into a triplet, with energies of $-3ea_0\mathcal{E}$, 0, and $+3ea_0\mathcal{E}$ with respect to the unperturbed level. As expected, the splitting is linear in the field.

Appendix E

Interaction with narrow-band radiation

The Einstein B coefficients were introduced in Chapter 9 to consider the interaction of atoms with broad-band radiation, such as black-body radiation, as illustrated in Fig. E.1(a). In this situation, the spectral energy density $u(\nu)$ varies much more slowly with frequency ν than the atomic lineshape function $g(\nu)$, and may effectively be taken as constant over the line width of the transition. In a laser, by contrast, the spectral width of the radiation inside the cavity is frequently much narrower than the width of the atomic transition, as illustrated in Fig. E.1(b).

Th absorption and stimulated emission transition rates for the case of narrow-band radiation, as shown in Fig. E.1(b), can be calculated as follows. The spectral line-shape function $g(\nu)\mathrm{d}\nu$ gives the probability that a particular atom will absorb or emit in the spectral range $\nu \to \nu + \mathrm{d}\nu$. Hence the number of atoms in the lower level per unit volume that can absorb radiation in this frequency range is $N_1 g(\nu)\mathrm{d}\nu$. From the definition of the Einstein B_{12} coefficient given in eqn 9.7, the absorption rate in this frequency range is therefore:

$$\mathrm{d}W_{12} = B_{12}\, N_1 g(\nu)\mathrm{d}\nu\, u(\nu)\,. \tag{E.1}$$

The total absorption rate is thus:

$$W_{12} = \int_0^\infty B_{12} N_1 g(\nu) u(\nu)\, \mathrm{d}\nu\,. \tag{E.2}$$

Since the spectral energy density of the radiation inside the laser cavity is much narrower than the width of the atomic transition, we can write it as:

$$u(\nu) = u_\nu \delta(\nu - \nu_{\mathrm{laser}})\,, \tag{E.3}$$

where u_ν is the total energy density of the beam (cf eqn 9.20) and $\delta(\nu)$ is the Dirac delta function. The Dirac delta function $\delta(x - x_0)$ takes the value of 0 at all values of x apart from x_0, and is normalized such that $\int_0^\infty \delta(x - x_0)\mathrm{d}x = 1$. It can be thought of as the limit of a top-hat function of width Δ and

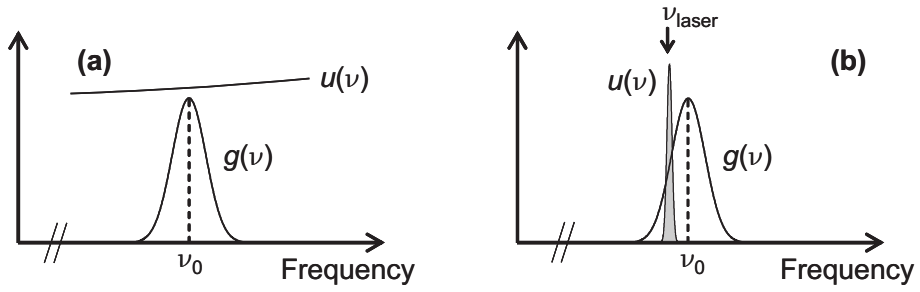

Figure E.1: Interaction of an atomic transition with: (a) broad-band radiation, and (b) narrow-band radiation. Note that the spectral energy densities and the atomic line-shape functions are not drawn on the same vertical scales.

height $1/\Delta$ centred at x_0 in the limit where $\Delta \to 0$. It is easy to show that

$$\int_0^\infty f(x)\delta(x - x_0)\,\mathrm{d}x = f(x_0)\,.$$

On inserting eqn E.3 into E.2, we obtain:

$$\begin{aligned} W_{12} &= \int_0^\infty B_{12}N_1 g(\nu)u_\nu \delta(\nu - \nu_{\mathrm{laser}})\,\mathrm{d}\nu\,. \\ &= B_{12}N_1 g(\nu_{\mathrm{laser}})u_\nu\,. \end{aligned} \tag{E.4}$$

The argument for the stimulated emission rate follows similarly, and leads to:

$$W_{21} = B_{21}N_2 g(\nu_{\mathrm{laser}})u_\nu\,. \tag{E.5}$$

Appendix F

Mathematics of mode-locking

The electric field of the light emitted by a multi-mode laser is given by:

$$\mathcal{E}(t) = \sum_m \mathcal{E}_m \exp(\mathrm{i}\omega_m t + \varphi_m)\,, \tag{F.1}$$

where the sum is over all the longitudinal modes that are oscillating. ω_m is the angular frequency of the mth mode ($= m\pi c/L$, for $n = 1$), and φ_m is its optical phase. In multi-mode operation all the phases of the modes are random, and not much can be done with the summation. However, in a mode-locked laser, all the phases are the same (call it φ_0) because they have been locked together. This allows us to evaluate the summation.

We assume that all the modes have approximately equal amplitudes $\mathcal{E}_0$. The output field of the mode-locked laser is then given by:

$$\mathcal{E}(t) = \mathcal{E}_0 \mathrm{e}^{\mathrm{i}\varphi_0} \sum_m \mathrm{e}^{\mathrm{i}\omega_m t}\,. \tag{F.2}$$

Let us suppose that there are N modes oscillating, and the frequency of the middle mode is ω_0. This gives the field as :

$$\begin{aligned} \mathcal{E}(t) &= \mathcal{E}_0 \mathrm{e}^{\mathrm{i}\varphi_0} \sum_{m'=-(N-1)/2}^{m'=+(N-1)/2} \exp\left(\mathrm{i}(\omega_0 + m'\pi c/L)t\right)\,, \\ &= \mathcal{E}_0 \mathrm{e}^{\mathrm{i}\varphi_0} \mathrm{e}^{\mathrm{i}\omega_0 t} \sum_{m'=-(N-1)/2}^{m'=+(N-1)/2} \exp\left(\mathrm{i}\frac{m'\pi c}{L}t\right)\,. \end{aligned} \tag{F.3}$$

This type of summation is frequently found in the theory of diffraction gratings. It can be evaluated by standard techniques.[1] We are actually interested in the time dependence of the output power, which is given by:

$$P(t) \propto \mathcal{E}(t)\mathcal{E}(t)^*\,. \tag{F.4}$$

The final answer is:

$$P(t) \propto \frac{\sin^2(N\pi ct/2L)}{\sin^2(\pi ct/2L)}\,. \tag{F.5}$$

This function has big peaks whenever $t = \text{integer} \times 2L/c$ and is small at all other times. Thus the output consists of pulses separated in time by $2L/c$. The duration of the pulse is approximately given by the time for the numerator to go to zero after one of the major peaks. This time is $2L/Nc$. The frequency band width $\Delta\nu$ is equal to the (number of modes oscillating) $\times$ (spacing between modes), i.e. $\Delta\nu \sim N \times c/2L$. Thus $\Delta t \Delta\nu \sim 1$ as expected from the uncertainty principle given in eqn 10.7.

[1] Remember that $\mathrm{e}^{ab} = (\mathrm{e}^a)^b$, and that $\sum_{j=0}^{n-1} r^j = (1 - r^n)/(1 - r)$.

Appendix G

Frequency conversion by nonlinear optics

It was discovered very early on in the history of lasers that certain crystals could double the frequency of laser light, as shown in Fig. G.1(a). This effect, which is know as **frequency doubling**, works by combining two photons at frequency ω to produce a single photon at frequency 2ω, as shown in Fig. G.1(b). It occurs when a **nonlinear crystal** is driven by the intense light field produced by a powerful laser. The crystal must be non-centro-symmetric: i.e., belong to a crystal class that does not have inversion symmetry. Beta-barium borate (BBO), potassium dihydrogen phosphate (KDP) and lithium niobate are well-known examples of such crystals.

Frequency doubling is a specific example of a more general effect called **nonlinear frequency mixing**. In nonlinear frequency mixing, the nonlinear crystal is driven by two intense waves at angular frequencies ω_1 and ω_2, and a third wave is generated at the sum of their frequencies, as shown in Fig G.1(c):

$$\omega_3 = \omega_1 + \omega_2 \,. \tag{G.1}$$

Frequency doubling corresponds to the case where $\omega_1 = \omega_2$. The condition in eqn G.1 is equivalent to energy conservation in the photon conversion process. The photon momentum must also be conserved, as shown in Fig. G.1(d), which requires that:

$$\boldsymbol{k}_3 = \boldsymbol{k}_1 + \boldsymbol{k}_2 \,, \tag{G.2}$$

where $\boldsymbol{k}_3$, $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ are the respective wave vectors inside the crystal. This condition is called **phase matching**, and is, in general, very hard to satisfy. Nonlinear frequency mixing therefore only works efficiently for the very specific wavelengths that satisfy the phase-matching condition. These wavelengths are selected by the orientation of the crystal.

The nonlinear process works equally well in reverse, and a single "pump" photon can be split into two photons of lower frequency called the "signal" and "idler" photons subject to energy conservation:

$$\omega_\mathrm{p} = \omega_\mathrm{s} + \omega_\mathrm{i} \,. \tag{G.3}$$

This process is called **parametric down conversion** and is illustrated in Fig. G.1(e). Note that there is an infinite number of combinations of signal and idler frequencies that can satisfy eqn G.3, and the actual frequencies that are produced are determined by the phase-matching condition. Down conversion is a convenient way to generate tuneable radiation from a fixed-frequency laser, and is now widely used to extend the range of frequencies available from lasers.

It should be pointed out that the frequency conversion processes that are considered in this appendix are examples of phenomena that are well-known in classical nonlinear optics. The description in terms of photons is helpful, but not necessary: all of the effects can have classical explanations. Quantum effects do show up when these nonlinear processes are considered at the single-photon level, but this is not the regime that is being considered when the driving field is an intense laser beam.

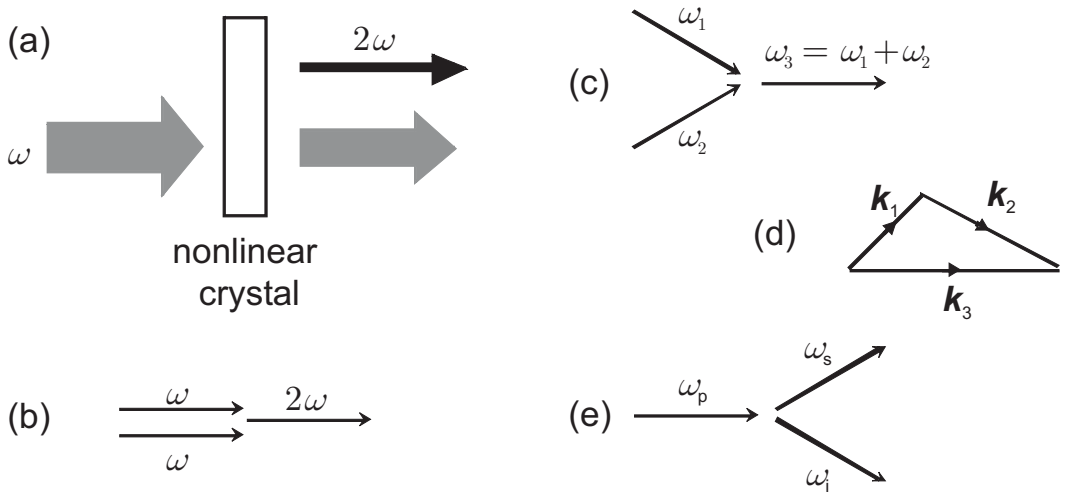

Figure G.1: (a) and (b): frequency doubling. (c) Sum frequency mixing. (d) Phase matching. (e) Parametric down conversion. The subscripts p, s, and i stand, respectively, for "pump", "signal" and "idler".