

Physics III

J. P. Home

September 23, 2012

Contents

1 Optics and optical instruments	7
1.1 Geometrical optics	8
1.1.1 Fermat's Principle	8
1.1.2 The lens	9
1.1.3 Ray tracing and image formation	13
1.2 Waves and optics	17
1.2.1 Superposition of waves	17
1.2.2 Huygens Diffraction Theory	18
1.2.3 Kirchoff Diffraction Theory	19
1.2.4 Fraunhofer diffraction	22
1.2.5 Huygen's principle and refraction	25
1.3 Spatial Resolution	26
1.3.1 Telescopes and the Rayleigh criterion	26
1.3.2 Microscopy and the Abbé criterion	27
1.4 Polarization	28
1.4.1 Types of polarization	28
1.4.2 Birefringence	29
1.4.3 Fresnel equations	30
1.5 Spectral measurements of light	34
1.5.1 The Prism	34
1.5.2 The Grating Spectrometer	34
1.5.3 The Michelson Interferometer	37

1.5.4	The Fabry-Perot Etalon	38
1.5.5	The use of etalons and gratings in Lasers	43
1.5.6	LIGO and LISA	44
2	Photons	45
2.1	The Photo-electric effect	45
2.1.1	Einstein's explanation	48
2.1.2	Photon Detection	48
2.2	X-Rays and Bragg scattering	50
2.2.1	Production of X-rays	50
2.2.2	Rayleigh scattering	52
2.2.3	Bragg scattering	53
2.3	Photon momenta – The Compton effect	56
3	Atoms and their composition	59
3.1	Introduction to atoms	59
3.1.1	Periodic table of elements	59
3.2	Measurements of atomic mass	62
3.2.1	Mass spectrometry	63
3.2.2	Isotopes	65
3.3	Size of the atom - scattering experiments	66
3.3.1	Estimating atom size by other methods	67
3.3.2	Comment on the size of the atom	68
3.4	Evidence for atomic structure	69
3.4.1	Electron scattering	69
3.4.2	The nucleus - Rutherford scattering	70
3.4.3	Electronic structure	73
4	Statistical Mechanics	77
4.1	Books and problems	77

4.2	Equilibrium statistical mechanics	77
4.2.1	Coin example	78
4.2.2	Continuous and discrete probability distributions	78
4.3	Kinetic theory and the ideal gas	79
4.4	The Boltzmann distribution	82
4.4.1	Coin flipping and the equilibrium state.	82
4.4.2	An isolated system of distinguishable particles	84
4.4.3	A system in thermal equilibrium with a heat bath	85
4.4.4	Example of use of the partition function	88
4.5	The equipartition theorem	89
4.5.1	The equipartition theorem for molecules	90
4.5.2	Mean energy for a discrete system	91
4.6	Blackbody experiments and concepts	93
4.6.1	Optical modes of a cavity	94
4.6.2	Classical and quantum distributions	95
4.6.3	Einstein and the Planck Distribution	97
4.7	Indistinguishable particles	99
4.7.1	Review: relation between partition functions for distinguishable particles	99
4.7.2	Counting states of indistinguishable particles	100
4.7.3	The classical limit - single particle partition function	101
4.7.4	The many-particle partition function	102
4.7.5	Back to the ideal classical gas.	103
4.7.6	The Maxwell-Boltzmann velocity distribution	104
5	Wave-particle “duality”	107
5.1	Matter waves	107
5.1.1	The Davisson-Germer experiment	107
5.1.2	de Broglie hypothesis	107
5.1.3	Modern applications.	108

5.2	What is a matter wave?	111
5.2.1	The particle as a wavepacket	111
5.2.2	Interpretation of matter waves	112
5.2.3	Diffraction and the uncertainty relation	113
5.2.4	Double-slit interference with matter waves	114
6	Quantum Mechanics	116
6.1	The Wavefunction	118
6.1.1	Expansion of functions	118
6.2	Physical variables as operators	122
6.2.1	Change of basis: position and momentum	123
6.3	Measurement	125
6.3.1	Example: position and momentum	125
6.3.2	Simultaneous sets of eigenstates	126
6.3.3	Expectation values	128
6.4	The Schrödinger equation	129
6.4.1	The free particle in 1-D	130
6.5	The Infinite Square Well	130
6.6	The Finite Square Well	134
6.6.1	Parity inversion	134
6.6.2	Continuity of the wave function	135
6.6.3	Solutions of the TISE	136
6.7	Tunneling	141
6.8	The Quantum Harmonic Oscillator	145
6.8.1	Ladder operators	146
6.8.2	Eigenstates	146
6.8.3	The ground state of energy	147
6.8.4	Energy levels	148
6.8.5	The excited states	149

6.8.6	Expectation values of the harmonic oscillator	151
6.8.7	Classical/Quantum correspondence	153
7	Hydrogen	154
7.1	Pre-quantum models	154
7.1.1	Early Hydrogen spectra	154
7.1.2	The Bohr model	156
7.2	Quantum mechanical treatment	159
7.2.1	Separation of variables	159
7.2.2	Angular momentum eigenfunctions	160
7.2.3	The radial solutions	165
7.2.4	Degeneracy of the Hydrogen wave functions	170
7.3	First approximations	171
7.4	The Normal Zeeman effect	174
7.4.1	The dipole moment of the orbiting electron	174
7.4.2	Interaction with the magnetic field	174
7.4.3	Magnetic field aligned along x - the problem of degeneracy	175
7.5	Optical transitions.	177
7.5.1	Linear polarization aligned with an external magnetic field	178
7.5.2	Circular polarization	180
8	Spin	182
8.1	Experimental observations	182
8.1.1	Doublets in spectroscopy	182
8.1.2	Stern-Gerlach experiment	182
8.1.3	Hypothesis of the electron spin	184
8.2	The quantum rules for spin	184
8.2.1	Matrix mechanics	186
8.2.2	Combining spin with the rest of physics	186
8.3	The intrinsic magnetic moment	187

8.3.1	The spin of other particles	188
8.4	The spin-orbit interaction	189
8.4.1	Perturbation Hamiltonian: semi-classical derivation	189
8.4.2	The spin-orbit energy level shifts	190
8.5	Hydrogen fine structure	195
8.5.1	Combined results.	195
8.5.2	The Lamb shift	195
9	Atomic and molecular physics	197
9.1	Helium	198
9.1.1	Exchange symmetry	198
9.2	Quantum statistics	201
9.2.1	Bose-Einstein distribution	201
9.2.2	Fermi-Dirac distribution	202
9.3	Many-Electron atoms	204
9.3.1	The central field approximation - configurations	204
9.3.2	L-S coupling	206
9.3.3	Spin-orbit coupling	207
9.3.4	$j - j$ coupling	208
9.4	The Periodic Table	209
9.4.1	The alkali group	209
9.4.2	The inert gases	210
9.5	Adding angular momenta	211
9.6	Optical transitions in many-electron atoms	212
9.6.1	The configuration	212
9.6.2	The term	212
9.6.3	The level	213

Chapter 1

Optics and optical instruments

We study optics because light (and beyond visible light, the whole electromagnetic spectrum) is the primary resource which allows us to investigate the physical world. This is not just true in physics - light is extremely important in biology, chemistry, crossing the road, reading a book etc..

The part of the course on light will introduce two different uses of light. The first is in imaging, where we make use of the position dependence of light. Here one of the most important ingredients is the lens, which allows us to collect and separate light from different sources efficiently, giving us good signal to noise for our measurements. Most work with lenses falls within the realm of *geometrical optics*, where the wavelength of the light does not play a strong role.

Some of the most important limits to the performance of imaging instruments are due to effects which do depend on the wavelength of light. Here, we must consider the effects of interference, in particular diffraction. However these properties open up new opportunities to investigate the spectrum of light (how much intensity appears at each color), providing a whole new field of measurement science. Measurements of spectra were the primary observations which led to the need for a new theory of physics in the 19th century. Spectral measurements are used throughout science today.

Interference and diffraction have now become tools for spatial imaging in their own right. Of particular note is the use of high frequency electromagnetic radiation (X-rays) for finding the structure of crystals, or molecules. A notable example of the use of interference for molecular structure is the deduction of the structure of DNA. We'll see examples of this type of science later in the course, building upon the work we cover here in optics.

So light is useful for measurement. On a more fundamental level, the methods developed for dealing with the propagation of light waves are a great precursor to the study of quantum mechanics. As we shall find out later in the course, quantum mechanics tells us that all matter can be viewed in a wave-particle picture, where interference and the phase of the quantum mechanical system are very important. Therefore we should also view optics as a building block towards understanding the quantum.

1.1 Geometrical optics

1.1.1 Fermat's Principle

Fermat's principle is one of the earliest rules specifying the path light takes between two points. Two statements of Fermat's principle are:

- Light passing between two points will take the path which is quickest.
- Light will take the path which is extremal with respect to variations in the optical path length.

You see already why people state it in different ways. The first statement is very straightforward, and doesn't require any knowledge of the type of thinking we do in a degree level physics course. The second requires us to think harder. What do we mean by *extremal* with respect to variations in the optical path? And what is the the *optical path length*? The optical path length is a measure of the time it takes for light to pass between two points. However the units are spatial. To convert the time taken to the optical path we multiply by c . For light passing between two points a distance s apart, the time taken is

$$t = \frac{s}{c} \quad (1.1)$$

if the region between the two points is filled by vacuum. The optical path length is

$$ct = s . \quad (1.2)$$

Now consider what happens if the light travels through a medium of refractive index n . Here the speed of light is c/n , so the time taken to travel a geometric distance s is

$$t' = \frac{ns}{c} \quad (1.3)$$

and the optical path is thus

$$D = ct' = ns . \quad (1.4)$$

Why is the time, written in terms of the optical path, an important feature of the light? Here it is worth thinking in terms of a wave picture of light. One quantity which is fixed whether the light is inside and outside of a material is the frequency of the light. Thus both the optical path and the time give us information about how many cycles of oscillation were completed in traveling from one point or another - this is information regarding the *phase* of the light. To obtain the number of oscillation cycles from the total time, we divide the time taken by the cycle time $t_c = 1/f$. We thus get

$$\frac{t'}{t_c} = \frac{fns}{c} = \frac{D}{\lambda_0} \quad (1.5)$$

where $\lambda_0 = c/f$ is the *vacuum wavelength* of the light.

One consequence of Fermat's principle is that in a uniform medium, light will travel in a straight line. This is the minimum optical path, and is thus an extremal point. Here the simple statement of Fermat's principle is good enough - the light gets from one point to another in the shortest time.

Fermat's principle refers to an extreme path. This means that if we deviate from the path "chosen" by the light by a small amount the total distance traveled along the path would *not*

have changed to first order (at a minimum or maximum, the first order term - the gradient - is zero). Think about your work on interference from last term. If light travels down multiple paths, but each path has the same number of cycles of oscillation, it constructively interferes. If it constructively interferes then we see high intensity. We can view Fermat's theorem as saying that the reason light travels in straight lines is because it explores all routes, but the only ones which interfere constructively are where there is constructive interference, namely where many routes have the same optical path. This happens at extrema with respect to the optical path. This argument should become clearer as we proceed through the course. Note that this is a wave argument which arrives at a particle-like result. Energy moves between two points in a straight line.

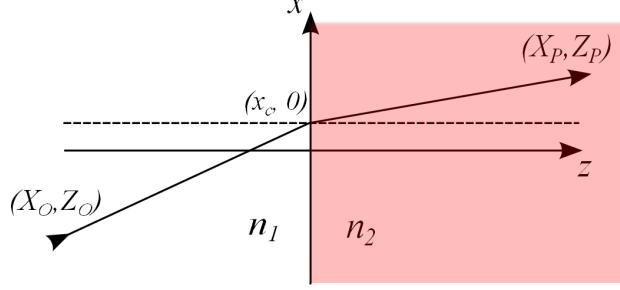


Figure 1.1: Refraction. Light propagating from a medium with refractive index n_1 into a second medium with refractive index n_2 . The path taken from (X_0, Z_0) to (X_P, Z_P) passes through $(x_c, 0)$, which can be found using Fermat's principle.

To further illustrate the use of Fermat's principle, let's consider a situation you've seen before, last term. What path does light take between two points, one which sits in one medium, the second in another? In the first material, the speed of the light is c/n_1 , and in the second it is c/n_2 . So the number of oscillations of the light in the two materials is $n_1\omega D_1/c$, and $n_2\omega D_2/c$, where D_1, D_2 are the spatial distances traveled in each material. Using the variables as defined in figure 1.1, the number of oscillations is

$$\frac{\omega}{c} \left(n_1 \sqrt{(x - X_0)^2 + Z_0^2} + n_2 \sqrt{(x - X_P)^2 + Z_P^2} \right) \quad (1.6)$$

We can find the minimum by differentiating with respect to x (the position at which the light crosses the surface is the only variable parameter), and solving. This gives

$$n_1 \frac{x - X_0}{\sqrt{(x - X_0)^2 + Z_0^2}} + n_2 \frac{x - X_P}{\sqrt{(x - X_P)^2 + Z_P^2}} = 0 \quad (1.7)$$

which is exactly the result you got last term, $n_1 \sin \theta_i = n_2 \sin \theta_j$: Snell's law.

A second example of Fermat's principle will be covered in the problem sheet, where you will use the same method to find the angle at which light reflects from a mirror.

One thing to note about Fermat's principle, especially in this course, because we will do quantum mechanics later, is that Fermat's principle seems to describe propagation in a similar way to how we talk about particles (it tells us that light travels in straight lines, for instance). In the context of particles, we ask ourselves: How does light know it should take the off-axis path through the first medium before it arrives at the second?

1.1.2 The lens

Among the most important optical elements is the lens. The most popular implementation for teaching purposes is the infinitesimally thin lens (something a theoretical physicist uses),

an approximate form which gives most of the correct results. Of course, no lens is strictly “thin”, so we might want to have a little understanding of the approximations made in coming to this form. Therefore in the next section we will try to find what shape a lens should be.

The ideal thin lens has the following action. A beam traveling along a straight line which hits the lens at a distance h from its center will be deflected by an angle $\arctan(h/f)$, where f is known as the focal length. Very often with lenses, we work in the limit that $h \ll f$, and that the light rays propagate at a small angle to the central axis, which we call the optical axis. This means that we can make the approximation $\arctan(h/f) \sim h/f$ (the same holds for other relations too). This is called making the *paraxial approximation*. The important results which you need to know in order to calculate the propagation of light through a thin lens are

- If a ray of light passes through the center of the ideal lens it will not be deflected.
- parallel rays hitting a thin lens will no longer be parallel, and will meet at a point a distance f behind the lens - this is the focal point.

These are very useful results, because they allow us to figure out where *images* of objects will appear. An image is the point where all rays emanating from a particular point on the object meet again. This is important because it provides a lot of signal from this one point on the source, relative to noise due to all the junk light that is flying around. This is a recurring theme in measurement science - the need to maximize signal compared to the background noise.

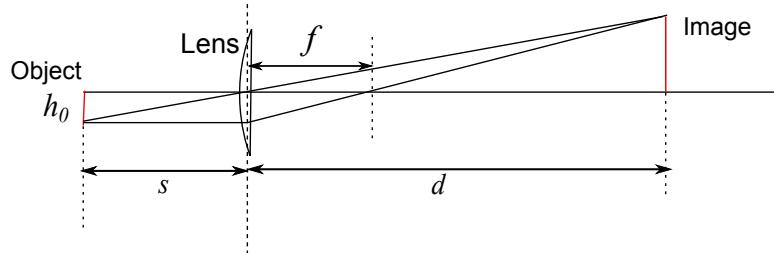


Figure 1.2: Reconstructing image formation by a lens. One ray passes from the point on the object through the center of the lens - it is therefore undeflected. The second ray travels towards the lens parallel to the optical axis. It is therefore deflected by the lens to pass through the focal point on the optical axis. Where these two rays meet an image of the point will form. The plane in which images form is called the *image plane*.

An example of finding the position of an image is shown in figure 1.2. We start by drawing a ray which travels from the point on the object furthest from the optical axis and passes through the center of the lens (and is therefore not deflected). We also take a second ray from the same point which travels perpendicular to the lens, which will then be deflected to pass through the focal point. Using the object height h_0 and distance from the lens s , the angle of the undeflected ray is h_0/s , where again I have made the paraxial approximation in order to avoid considering the tangent. The perpendicular ray is deflected such that it propagates at an angle h_0/f to the optical axis. Thus the two beams will meet at a position beyond the lens d which satisfies

$$(s + d) \frac{h_0}{s} = d \frac{h_0}{f} \quad (1.8)$$

giving the famous thin-lens formula relating the position of objects and images to the lens

focal length

$$\frac{1}{s} + \frac{1}{d} = \frac{1}{f} \quad (1.9)$$

You will use this formula in the exercises.

A final comment about the action of lenses. The ideal lens transforms information specifying the position of a ray into information about the angle of the ray. The most important special case of this is for light coming from points a distance f from the lens. These end up forming an image at infinity (where is infinity?). A given point in the focal plane gives parallel light which propagates at a given angle after the lens. The converse is also true. All the light which is incident on the lens at a given angle will meet at a single point in the focal plane. This is very important in diffraction (see below), which typically causes angular divergence of light. In order to observe diffraction, we commonly want to take all light from a given direction and concentrate it at a particular position on our detector (the back of our eye, or the CCD on a camera). To do this, we generally use a lens, placing our detector in the focal plane.

The single-surface lens

How do we produce the best lens? The first thing to note is that a lens is designed for a particular job. If we are trying to image a point 4 cm in front of a lens to an image 30 cm behind, that requires a different lens to trying to image a point very far away to an image close behind. So in designing a lens, we must first ask what it is being used for. For our example below, the job of the lens is to take light from a distant star, and focus it down at a focal plane a distance f behind the lens. Since the star is so far away, we can assume that the light hitting different parts of the lens is all parallel (the range of straight lines you can draw from any point on the star to any point on the lens have very small differences in angle).

Let us consider Fermat's principle and see if we can derive a suitable form. To simplify matters, we make the simplest lens by considering a boundary between two materials, one with refractive index n_1 and the other with refractive index n_2 , as shown in figure 1.3. In

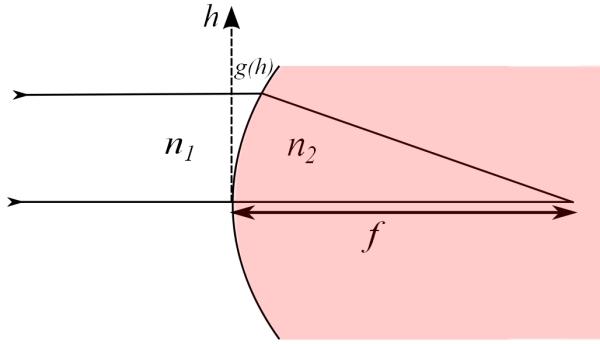


Figure 1.3: A surface between two materials of different refractive index.

this case, $n_2 > n_1$. The lens will be produced due to the fact that the boundary is not flat. From Fermat's principle and the condition which we imposed that rays parallel to each other and to the optical axis must come to a focus at the focal point, we know that these parallel rays must all have the same optical path. In distance, those which start further from the optical axis (which pass through the center of the lens) must travel further. Therefore we

need to slow down the rays in the middle relative to those on the outside. We can do this by making them travel for longer through the “slow” material n_2 . The total optical path taken for the ray starting at height h is

$$n_1g(h) + n_2(h^2 + (f - g(h))^2)^{1/2} \quad (1.10)$$

if f is the distance from the position of the lensing surface on the optical axis to the point at which the beams come to a focus. $g(h)$ is the function which governs the extra path through material n_1 for light hitting the boundary at height h . According to Fermat, we must find the stationary point with respect to changes in the optical path. In our case, the optical path is different depending on the height of the initial ray, so we differentiate with respect to h and set this equal to zero. This gives

$$n_1g'(h) + \frac{n_2(2h - 2(f - g(h))g'(h))}{2(h^2 + (f - g(h))^2)^{1/2}} = 0 \quad (1.11)$$

I put this there just so you can see that this problem is more tricky than you might suspect: the multiple appearances of $g(h)$ and $g'(h)$ make this a rather challenging analytical problem. It is also extremely tricky to fabricate - consider machining something to within a tolerance of a wavelength of light, but which is not a simple shape! I'll provide a series solution of an approximate solution to this formula later on, but let us first consider a surface which we can analyze and fabricate.

Note that since the lens is symmetric about the optical axis, it should be composed of even powers of h . Usually we use lenses in a form where most of the light propagates at a small angle to the optical axis and stays close to this axis, so let us consider an expansion of $g(h)$ around $h = 0$. The expansion of a general function $g(h) = \sum_n a_{2n}h^{2n}$ up to fourth order turns equation 1.10 into

$$fn_2 + \frac{(n_2 + 2f(n_1 - n_2)a_2)h^2}{2f} + O[h^4] \quad (1.12)$$

It is easy to see that to cancel the dependence of this distance to second order in h , we need

$$a_2 = -\frac{n_2}{(2(n_1 - n_2)f)} \quad (1.13)$$

What type of surface would that require? Since it only contains h^2 terms, it is clear that this is a parabola, so it seems like a parabolic surface would do the job. It turns out parabolic surfaces aren't that easy to fabricate either, so usually a spherical surface is used instead. For a spherical surface with radius of curvature $R \gg h$, we can write

$$g(h) = R - \sqrt{(R^2 - h^2)} \approx \frac{h^2}{2R} + O[h^4]. \quad (1.14)$$

Therefore we are looking for a radius of curvature for which $a_2 = 1/(2R)$, which along with equation 1.13 gives

$$R = \frac{f(n_2 - n_1)}{n_2} \quad (1.15)$$

or

$$f = \frac{R}{\left(1 - \frac{n_1}{n_2}\right)}. \quad (1.16)$$

In order to make a strong lens (short focal length) we therefore need a small radius of curvature. This means that the lens surface has high curvature - and in general requires the lens to be thick.

Let's go back to equation 1.11 to illustrate an important point about real lenses. We made an approximation in ignoring the higher order terms, and then another in approximating a

parabolic surface with a spherical surface. What is the leading correction in all this? Using the expansion of the spherical surface to fourth order, and putting this back into equation 1.11 we find that the leading error in our lens is

$$\frac{n_2 + \frac{2n_2^2}{n_1 - n_2} + \frac{(n_1 - n_2)(n_2 - n_1)^3}{n_2^2}}{8} \frac{h^3}{f^3} h \quad (1.17)$$

We call these errors *aberrations*. There are many different ways in which aberrations occur. The ones you see here are called *spherical* aberrations, and are present because the surface shape is not that which would satisfy Fermat's principle exactly. One other type are known as *chromatic* aberrations, because n_2 and n_1 typically depend on the wavelength. Finally, it is worth noting that in the derivation above, I considered light which is focussed on the optical axis. A more careful lens design would consider light incident at an angle to the axis, which is focussed to a point off-axis. All these points make lens design a pretty complex issue, and in many cases the use of the lens must be known before a design can be made. However for many cases of practical interest we don't need to worry too much. Why?

Note the way in which I wrote the dependence on h in the expression above. There is a linear scaling, which would be expected (as we scale up the whole system, we'd expect the error in a length to scale proportionally). However there is also a component which depends on the ratio $(h/f)^3$. So as long as we are working with situations where $h \ll f$, these corrections are small. We'll find later on that for achieving high resolution in a microscope we want to use a large diameter lens, because of limits due to diffraction. We should bear in mind though that if the lens gets too large, we may be limited instead by the spherical aberrations, since they scale rapidly with h/f .

The working described above is only part of the story, for instance real lenses tend to have two surfaces. There are also other methods of obtaining the form of a lens (ie. using Snell's law at each surface). These are all equivalent, but useful in different contexts. However now that you've got the general idea I'll leave you to learn this stuff through the problem sets.

1.1.3 Ray tracing and image formation

So we've dealt with a single surface lens, and we've discussed the basic function of a lens. However most measurement instruments consist of more than one lens, so we must also know how to deal with this situation. To do this, I will use several examples which should be familiar to you from your childhood.

Single-lens imaging instruments

Most humans make use of one optical instrument more than any other - the eye. This is a great instrument, but it has its restrictions. The elements of the eye are shown in figure 1.4. It consists of a lens of variable focal length, and a pixelated array of sensors on the retina, the back plane of the eye. This forms the basis of many imaging systems - your digital camera does something very similar (figure 1.4 b). Both the camera and the eye are restricted in several senses. The first is that the lens doesn't have infinite capacity to change in focal length - for the eye, this is the reason that we cannot read things closer than 15 cm or so (do the experiment, you'll "see" this). The second problem we have is that the retina (or in the camera the CCD) is at a fixed position, so we cannot aid the lens by moving it towards or away from the retina. Finally, the pixel array is fixed - the pixels have a finite size, and extend over a finite region on the back of the retina or CCD. These restrictions cause us problems in a range of different situations, which has led to the development of a range of optical instruments which modify incoming light so that we *can* create an image of

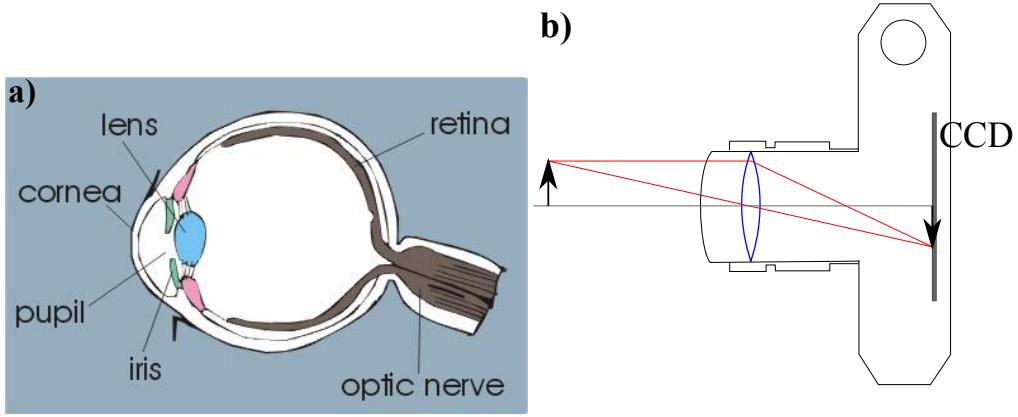


Figure 1.4: Simple imaging instruments. a) the eye, and b) the camera. Both have adjustable lenses which are used to focus light onto a sensor. For the eye this is the retina, while cameras these days tend to have CCD or CMOS sensors. Note that in both cases the image is inverted (for the eye, our brain is trained to invert the image). An important additional role is played by the iris, which controls the amount of light coming into the eye, so that the sensors don't saturate or vice versa. In a camera you control this with the shutter speed and an adjustable aperture (not shown).

objects which cannot be otherwise be imaged by the eye. Two of the most famous examples are introduced below.

The telescope

When do we use a telescope? Well generally we think in terms of looking at stars. But I can see many stars when I look at the sky? No, the reason we use a telescope is because we want to look at details in small regions of the sky, which means being able to separate light which comes from objects for which the angle subtended by the objects at the observation point is very small. If this happens, the image of both stars formed by our eye sit on a single pixel of the retina, and we miss out on the details. So somehow we have to magnify this angle, and this is what a telescope does.

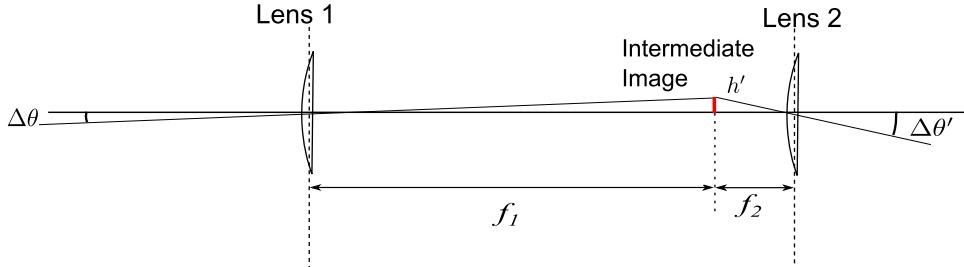


Figure 1.5: A telescope arrangement of lenses.

A common arrangement for a telescope is in figure 1.5. The incident light comes from two stars off left (we'd need a big page to put them on here), and the difference in angles which this light subtends at the beginning of the telescope is $\Delta\theta$. So the light passes through the first lens, and of course light passing through the center of the lens passes straight through. Let us assume that star 1 is on the optical axis of the system. Since all the light from star 1 comes from roughly the same angle, all of this light will form a focus (the image) at the focal point behind the lens. The same will be true for star 2, which will be offset from the optical axis by $h' = f_1 \tan(\Delta\theta)$. The second lens is placed such that light from each star

will be collimated after the lens, but this lens has a shorter focal length. The angle between the rays from the two stars after the lens is $\tan(\Delta\theta') = h'/f_2$. Substituting for h' we find that

$$\tan(\Delta\theta') = \frac{f_1}{f_2} \tan(\Delta\theta) \quad (1.18)$$

$$\Delta\theta' \approx \frac{f_1}{f_2} \Delta\theta \quad (1.19)$$

where the second expression is usually a good approximation, because the angles are generally small (otherwise we might not bother with the telescope!). Therefore the telescope can enhance the angular divergence by a factor f_1/f_2 , allowing the lens of our eye to image the two stars on separate pixels on our retina.

The microscope

Our eye is not able to focus on objects closer to it than a certain distance. This is because the angular divergence between light arriving from the object becomes too large, and our lens doesn't have the power to bend the rays in such that they can form an image on the retina, which is a fixed distance away. This causes a problem if we want to see very small details. To see why, consider trying to look at two points spaced by δ on an object a distance

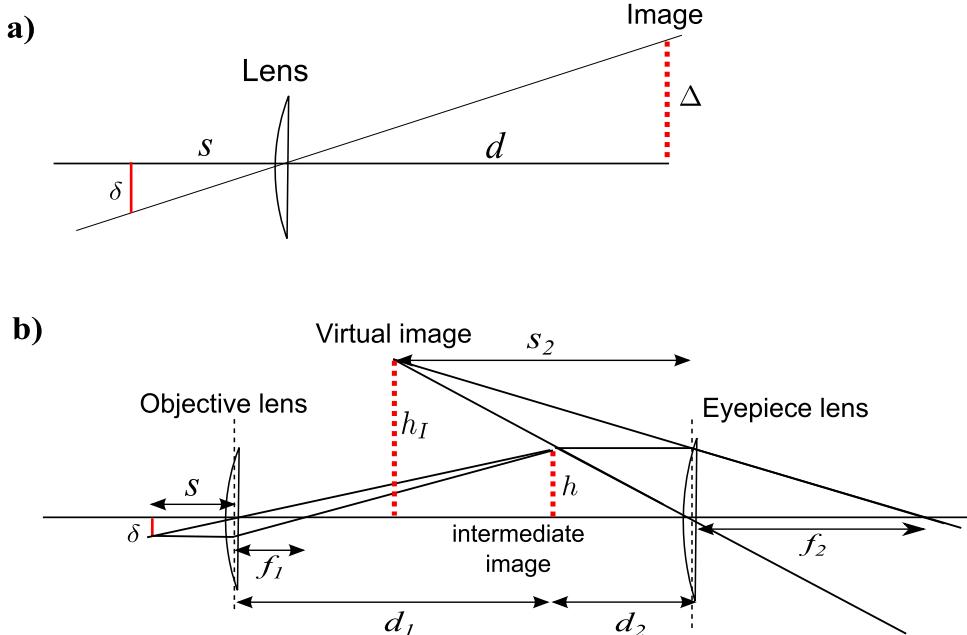


Figure 1.6: a) magnification of an object. If it is possible to form an image, then we get a magnified image if the object distance is smaller than the image distance. b) Microscope arrangement. Use of intermediate and virtual image allows a small close object to be viewed as though it were larger but further away. The eye is then able to image it.

s in front of the lens of our eye - see figure 1.6 a). The retina is a distance d behind the lens, which has a focal length f . If one of the objects is on the optical axis, and the other is not, we can see that if the lens can create an image on the retina, the size will be increased

$$\tan(\delta/s) = \tan(\Delta/d) \quad (1.20)$$

$$\frac{d}{s} \approx \frac{\Delta}{\delta} \quad (1.21)$$

So bringing the object closer to the eye will help us see small distances. However, we should also consider whether we can form an image, which requires us to satisfy the relation in

equation 1.9. We can see there that if d is fixed, decreasing s requires us to use a more powerful lens, which at some point our eye cannot provide.

In a microscope, we overcome this problem using the setup shown in figure 1.6. The object sits a short distance s from the first “objective” lens, which has a focal length $f_1 < s$. An image thus forms at a distance $1/d_1 = 1/f_1 - 1/s$ from the lens. The distance between the images of the two object points (again using rays from the two points on the object which pass undeflected through the center of the lens) is

$$h \simeq \delta \frac{d_1}{s} \quad (1.22)$$

This may already be enough to magnify the object, so that we could see it by eye (most of us did this as kids). However in general a microscope also uses a second lens, called an “eyepiece”. The eyepiece provides both extra magnification, and makes the microscope shorter for a given magnification. The eyepiece has a focal length f_2 and is placed a distance d_2 from the intermediate image, although now we arrange that $f_2 > d_2$. So the image forms at

$$\frac{1}{s_2} = \frac{1}{f_2} - \frac{1}{d_2} \quad (1.23)$$

which means that s' is negative! This means that the image forms to the left of the second lens, but how does the light get there? It doesn’t, but the light appears to come from a source point to the left of the lens, forming what we call a *virtual* image. The distance between the object points as they appear in this new image is given as before, by $h_I = hs_2/d_2$, giving a total magnification of $h_I/\delta = d_1s_2/(d_2s)$. Now the image is a) large, and b) far from the eye, so our eye is now able to see fine details.

1.2 Waves and optics

Viewed from electromagnetism, optics is all about the propagation of electromagnetic waves. So though in the previous section, we drew lots of lines alluding to the path of light (much as we would for particle trajectory), this should be consistent with the propagation of waves if we want physics to fit together. The use of waves also allows us to bring in the concepts of diffraction, which are extremely important. I'm aware that you learnt a lot about solutions to the wave equation last term, including using Huygen's principle for single and multi-slit experiments. Now, I want to approach this in a slightly more rigorous way, and show you when and where you can use "rays", as we did above, "Fraunhofer diffraction" which is the real name for what you did last term, and "Fresnel diffraction theory", which we will not discuss in any detail but is worth being aware of. The general problem which we consider

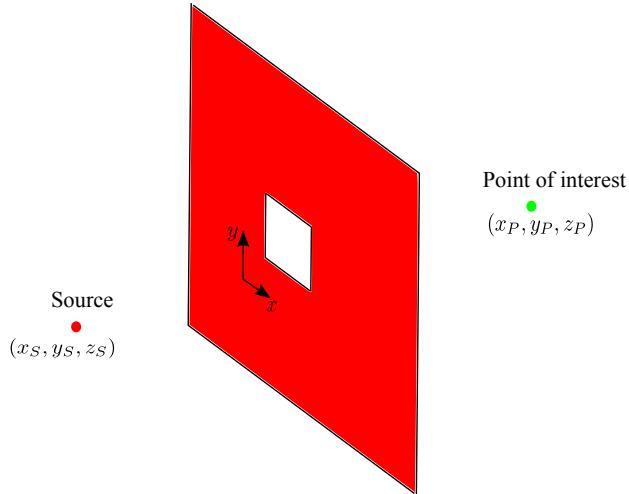


Figure 1.7: Typical problem in wave optics. Light propagates from a source at point S through a restricted aperture in a surface. We wish to find the resulting light field at the point P on the other side of the surface.

is the following. Given some source of electromagnetic radiation, and some restrictions on where the radiation can pass, what is the electromagnetic "disturbance" which occurs at another point in space? One example is shown in figure 1.7. There is a point source, and a restrictive aperture - this is the sort of thing you might remember considering last year. Any imaging system looks something like this - the sizes of the lens usually being the restrictive aperture. It is so important to measuring instruments across physics that we will look at it again now.

1.2.1 Superposition of waves

Wave optics is concerned with the propagations of electromagnetic waves, which are solutions of Maxwell's equation. It turns out that solving Maxwell's equations for wave optics problems is generally a very arduous task which is only truly solvable in a few special cases. Therefore we will start by simplifying the problem by considering scalar waves, as you did last term.

An important feature of the wave equation is that different solutions can be added together, and we still get a solution of the wave equation. This is the principle of *superposition of waves*. Most of what we will do for the rest of the part of the course concerned with optics is actually dealing with adding up waves, taking account of both the amplitude and phase of the summed constituents.

1.2.2 Huygens Diffraction Theory

We will next revise what you learnt about Huygens' principle next term, and try to use some intuition to write down the form of an equation which can describe what's going on. In the next section we give a more serious derivation of the same equation, which was arrived at by Kirchoff more than one-hundred years after the work of Huygens. Interested readers should take a look at this derivation (I think it is a example of a very elegant approach which is useful for many problems in electromagnetism), and I will cover it in a “special” class midway through term if people are interested. You do not need to know the Kirchoff derivation for the exam.

Huygens principle states that each element of a wavefront is a secondary source of a spherical wave, which has amplitude proportional to that of the original wavefront. The interference of these secondary waves produces the ongoing wavefront. You saw last year that this produces effects such as diffraction. The question now is how we write down Huygens principle mathematically?

Let us first consider that we are using Huygens principle, and write down a form for the secondary spherical wave propagating from a point on the wavefront. First, we note that the equation for an outgoing spherical wave from point x, y, z which has an amplitude A at this point is

$$A \frac{e^{i(kr - \omega t)}}{r} \quad (1.24)$$

where $r = \sqrt{(x_P - x)^2 + (y_P - y)^2 + (z_P - z)^2}$ and k, ω are the wavevector and angular frequency of the wave. The amplitude of the oscillations of this wave at any point in space x_P, y_P, z_P is thus

$$A \frac{e^{ikr}}{r} \quad (1.25)$$

The total emission from an infinitesimal area of a wavefront in the $x - y$ plane (ie. defined by the points $x, y, z, x + dx, y, z, x + dx, y + dy, z$ and $x, y + dy, z$) is

$$A(x, y, z) \frac{e^{ikr}}{r} dx dy \quad (1.26)$$

where the amplitude $A(x, y, z)$ is now written as a function of position. In writing this expression, I made the assumption that the amplitude of the wave doesn't change across this infinitesimal element of area.

In order to use Huygens principle to solve the problem depicted in figure 1.7, we now need a form for the amplitude of the wave which arrives at position x, y, z from the point source, which is situated at x_S, y_S, z_S . Since we have spherical symmetry about the source point, we might choose an outgoing spherical wave here too, for which the amplitude of the wave a distance $s = \sqrt{(x - x_S)^2 + (y - y_S)^2 + (z - z_S)^2}$ from the source is given by

$$A(x, y, z) = A_0 \frac{e^{iks}}{s} \quad (1.27)$$

where A_0 is the amplitude of the source.

Putting these two parts together, we find that the total amplitude arriving at x_P, y_P, z_P via the element of area dx, dy, dz is given by

$$A_0 \frac{e^{iks}}{s} \frac{e^{ikr}}{r} dx dy dz \quad (1.28)$$

which if we integrate over the aperture region (where the light can pass) gives a total amplitude of

$$U(P) = A_0 \int \int \frac{e^{iks}}{s} \frac{e^{ikr}}{r} dx dy . \quad (1.29)$$

If you compare this form with equation 1.47 from the next section, you will see that it has some similar features. However as might be expected there are additional factors which come from a more rigorous derivation, including additional constants and a factor which depends on the angle between the plane of the aperture and the lines connecting the source and destination positions to the center of the aperture. For our purposes here, we will primarily consider light propagating close to the normal to the plane of the aperture, and we won't look at absolute intensity, but relative values. This means that we don't need to worry about the extra constants and angular factors, and thus if you like you can jump to section 1.2.3.

1.2.3 Kirchoff Diffraction Theory

In order to get some useful formulae, we first need to look at a mathematical result for vector fields. These were derived by one of the most famous names in calculus, namely Green. What we will derive is Green's theorem. In fact, Green gave us many of the tools for solving complex electromagnetism problems, including functions very similar to the approach you will see here which provide nice routes to a number of electrostatic problems, but we don't have time to go into them all here (interested audience members can look them up in Jackson, Classical Electrodynamics).

Green's theorem

The result we will make use of comes from considering the divergence theorem

$$\int \nabla \cdot \mathbf{G} d^3r = \oint \mathbf{G} \cdot \mathbf{d}^2\mathbf{S} \quad (1.30)$$

where I use d^3r for an element of volume, $\mathbf{d}^2\mathbf{S}$ for an element of area (the direction is of course take normal to the surface) and $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$. \mathbf{G} is a **vector** field, which we can choose freely - so let us do that, and (because we know a useful result comes from this choice) use $\mathbf{G} = \phi \nabla \psi$, where ϕ and ψ are scalars. We find

$$\int (\nabla \phi \cdot \nabla \psi + \phi \nabla^2 \psi) d^3r = \oint (\phi \nabla \psi) \cdot \mathbf{d}^2\mathbf{S} \quad (1.31)$$

Repeating for the opposite choice of $\mathbf{G}' = \psi \nabla \phi$, and subtracting the two, we find

$$\int (\phi \nabla^2 \psi - \psi \nabla^2 \phi) d^3r = \oint [\phi(\nabla \psi) \cdot \hat{\mathbf{n}} - \psi(\nabla \phi) \cdot \hat{\mathbf{n}}] da \quad (1.32)$$

where $\hat{\mathbf{n}}$ is the outward unit normal to the surface. Let's take a breather and feel relaxed about the fact that the most abstract mathematical bit is over.

Spherical waves

Any pair of functions which obey Laplace's equation will make the left hand side of equation 1.32 go to zero. This could be of interest in electrostatics (it is!). Alternatively (and relevant here) we could look at solutions to the *wave* equation

$$\nabla^2 V - \frac{n^2}{c^2} \frac{\partial^2 V}{\partial t^2} = 0 . \quad (1.33)$$

Using a trial solution of the form

$$V(\mathbf{r}, t) = V(\mathbf{r})e^{i\omega t} \quad (1.34)$$

and performing the differential with respect to t , we obtain the Helmholtz equation

$$(\nabla^2 + k^2)V(\mathbf{r}) = 0 \quad (1.35)$$

where $k = \omega n/c$. If we now choose $\phi = U(\mathbf{r}), \psi = V(\mathbf{r})$ which are both solutions of the wave equation, then we find

$$\int (Uk^2V - Vk^2U) d^3r = 0 \quad (1.36)$$

and therefore

$$\oint [U(\nabla V) \cdot \hat{\mathbf{n}} - V(\nabla U) \cdot \hat{\mathbf{n}}] da = 0 \quad (1.37)$$

where $\hat{\mathbf{n}}$ is the unit normal to the surface element of area da .

Let us now remind ourselves of the conditions for which this equation is valid. Our derivation used the divergence theorem from electro-magnetism, which is valid for continuous functions with no poles (infinities). This means that if we are going to perform the integral over a closed surface, we should make sure that there are no poles within the volume enclosed by the surface. In the next step we are going to use spherical wave solutions to the wave equation, so we need to be careful that the centers of these waves are outside the integration volume.

As you found out last term, the solutions to the wave equation in spherical co-ordinates are

$$V(r, t) = \frac{a}{r}f(kr - \omega t) + \frac{b}{r}g(kr + \omega t) \quad (1.38)$$

Let us first look at incoming spherical waves, and imagine them to be impinging on the “destination” point P which lies at x_P, y_P, z_P , ie. the point at which we want to know the field. The solution of the wave equation in spherical coordinates gives as solution

$$V(r, t) = V_0 \frac{e^{ikr+i\omega t}}{r} \quad (1.39)$$

so the spatial part is $V(r) = e^{ikr}/r$ with $r = \sqrt{(x - x_P)^2 + (y - y_P)^2 + (z - z_P)^2}$. As we forewarned previously, there is a pole in this function at the point P where $r \rightarrow 0$ and thus $V(\mathbf{r}) \rightarrow \infty$. So if we use solutions like these, we had better enclose a volume which doesn’t include P . As a result of this consideration, we use a surface which is composed of two pieces. The inner piece is spherical and is centered on P with radius R . The outer surface can be of arbitrary shape, but the distance to P is always greater than the radius of the inner part. Note that for the inner surface, the normal points outwards from our volume towards point P , and since this surface is a sphere centered on P is in the opposite direction to r .

The function V is a scalar function with a scalar argument r . The gradient of this function is thus

$$\nabla V(r) = \frac{\partial V}{\partial r} \nabla r \quad (1.40)$$

$$= V_0 \left(ik - \frac{1}{r} \right) \frac{e^{ikr}}{r} \nabla r \quad (1.41)$$

$$= V_0 \left(ik - \frac{1}{r} \right) \frac{e^{ikr}}{r} \hat{\mathbf{r}} \quad (1.42)$$

On the inner (spherical) surface, $\hat{\mathbf{r}}$ is the opposite direction to $\hat{\mathbf{n}}$, and since it is a spherical surface at the surface element is $R^2 d\Omega$ where $d\Omega$ is the solid angle. For this surface we thus have an integral

$$\oint_{\text{inner}} U(R) \left(ik - \frac{1}{R} \right) \frac{e^{ikR}}{R} (-R^2 d\Omega) - \frac{e^{ikR}}{R} (\nabla U) \cdot \hat{\mathbf{r}} R^2 d\Omega \quad (1.43)$$

which as $R \rightarrow 0$ and taking the integral over the full 4π solid angle becomes $4\pi U(P)$ since all other terms will become negligible.

Putting this result back into 1.37, we find that

$$U(P) = -\frac{1}{4\pi} \oint_{\text{outer}} U(\mathbf{r}) \left(ik - \frac{1}{r} \right) \frac{e^{ikr}}{r} \hat{\mathbf{r}} \cdot \hat{\mathbf{n}} da + \frac{e^{ikr}}{r} (\nabla U(\mathbf{r})) \cdot \hat{\mathbf{n}} da \quad (1.44)$$

This is called the Kirchoff integral theorem. We shall not explore the general use of such a formula, but shall quickly move on to a point source. This source is the point S , situated outside the outer surface. Since this is a point source, we can imagine that the waves from this source are outgoing spherical waves, and therefore

$$U(s)e^{-i\omega t} = A \frac{e^{i(k_s s - \omega t)}}{s} \quad (1.45)$$

where $s = \sqrt{(x - x_S)^2 + (y - y_S)^2 + (z - z_S)^2}$. In evaluating the contribution from such a source, we'll have to perform the same differentiation as for equations 1.40 to 1.42, giving

$$\nabla (e^{iks}/s) \cdot \hat{\mathbf{n}} = A \left(ik - \frac{1}{s} \right) \frac{e^{iks}}{s} \hat{\mathbf{s}} \cdot \hat{\mathbf{n}} \quad (1.46)$$

We now make our first approximation. We say that $r, s \gg \lambda$, so that the ik terms in the differential dominate the $1/r$ terms. Putting this together with equation 1.44 we find that

$$U(P) = -\frac{iA}{4\pi} \oint_{\text{outer}} \frac{e^{ik(s+r)}}{rs} (k\hat{\mathbf{r}} \cdot \hat{\mathbf{n}} - k\hat{\mathbf{s}} \cdot \hat{\mathbf{n}}) da \quad (1.47)$$

This is the *Kirchoff-Fresnel diffraction* equation. It is the result which we will use to investigate diffraction, and we will show that it leads us in the correct limits to Huygen's principle, Fermat's principle, and as the name would suggest, Fraunhofer and Fresnel diffraction.

Simplification

How do we use this mathematical result? Remember that we have not yet described what the surface looks like, so we are free to choose a surface which fits our problem. So now let us consider this point. In particular, we will focus on the surface shown in figure 1.8, which is split into three parts. Parts A and B lie in the plane $z = 0$, with A transparent and B opaque. This means that for the part of the surface B , there is no disturbance to integrate over. Part C bounds the whole of space, and is at an infinite distance from the point P . We shall ignore it in what follows. You might worry about its contributions, but you'll have to trust me that they're small. A simple argument, but not quite valid for monochromatic light, is that if it is far enough away then the light approaching surfaces B and A from the left can not reach C before they reach P .

With this surface definition, we can now restrict our integral to the region A . Note that $e^{ik(r+s)}$ will oscillate over a full cycle if $r + s$ change by the wavelength λ . By contrast, $\hat{\mathbf{s}} \cdot \hat{\mathbf{n}}$ and $\hat{\mathbf{r}} \cdot \hat{\mathbf{n}}$ vary between -1 and 1 over all possible angles, and the $1/(rs)$ term only varies if the surface protrudes a significant fraction of the distance between either source and surface or surface and destination. The consequence is that we shall primarily be concerned with

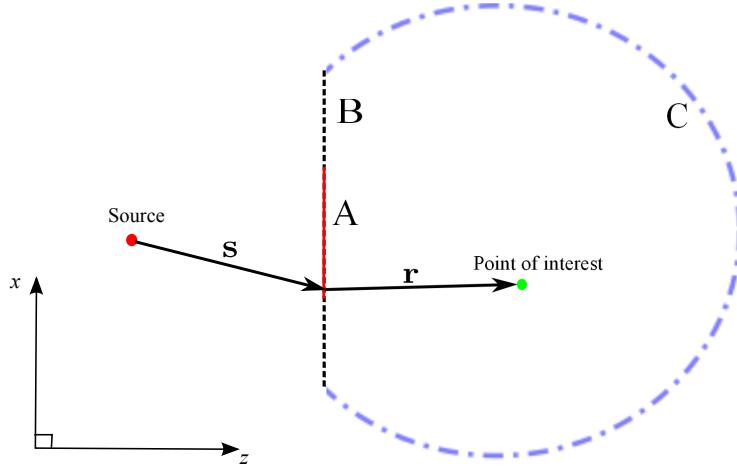


Figure 1.8: Cross-section view of splitting the outer closed surface into different regions in order to perform calculations of wave propagation through restricted apertures. Area A (red, solid) is the region which will be integrated over. B (dashed, black) is opaque, and thus no disturbance can pass through. The region C (dot-dashed, blue) is taken to infinity, which allows us to neglect its contribution.

the behavior of the exponential part of the integral, and shall largely ignore the rest. The result is that we work with a simplified integral

$$U(P) = \frac{iA}{4\pi r_0 s_0} (k\hat{\mathbf{r}}_0 \cdot \hat{\mathbf{n}} - k\hat{\mathbf{s}}_0 \cdot \hat{\mathbf{n}}) \int_A e^{ik(s+r)} da = K \int_A e^{ik(s+r)} da \quad (1.48)$$

where r_0, s_0 are the distances from the centre of the surface A to the points P and O respectively. We will examine the behavior of the integral in two different regimes, namely

- When λ is similar to the aperture size. This leads to diffraction.
- When λ is much smaller than the aperture size, but the aperture size is still much smaller than r, s . We recover geometrical optics in this limit, which we will show later by deriving Snell's law.

1.2.4 Fraunhofer diffraction

In this section we will use equations 1.47 and 1.48 to derive the diffraction of light through a finite aperture. We will work in the limit that $\lambda \sim \Delta_x, \Delta_y$ where Δ_x and Δ_y give the x and y extents of the area A . As a consequence, $\Delta_x, \Delta_y \ll r_0, s_0$ which allows us to expand the distances s and r about the values they hold at the center of the aperture r_0 and s_0 .

$$s = \sqrt{(x - x_S)^2 + (y - y_S)^2 + (z - z_S)^2} \simeq \left(s_0 - \left(\frac{xx_S}{s_0} + \frac{yy_S}{s_0} \right) + \frac{(x^2 + y^2)}{2s_0} \right) \quad (1.49)$$

$$r = \sqrt{(x - x_P)^2 + (y - y_P)^2 + (z - z_P)^2} \simeq \left(r_0 - \left(\frac{xx_P}{r_0} + \frac{yy_P}{r_0} \right) + \frac{(x^2 + y^2)}{2r_0} \right) \quad (1.50)$$

where higher order terms have been dropped since they will be small. If we go further, and drop the quadratic terms, we get a linear dependence, and the integral from equation 1.48 becomes

$$U(P) = K e^{ik(r_0 + s_0)} \int_{-\Delta_x/2}^{\Delta_x/2} \int_{-\Delta_y/2}^{\Delta_y/2} e^{-ik(Xx + Yy)} dx dy \quad (1.51)$$

where

$$X = \frac{x_S}{s_0} + \frac{x_P}{r_0} \quad (1.52)$$

$$Y = \frac{y_S}{s_0} + \frac{y_P}{r_0} \quad (1.53)$$

The approximation that we keep only linear terms in the exponential corresponds to the *Fraunhofer limit* of diffraction. But what exactly does this approximation mean? From the perspective of the spherical waves propagating from the source point, it means that the wavefronts are being approximated as plane waves across the area A . Note that if these are propagating at an angle to the slit, the phase of these waves will vary across the area, but this will be linear, because they are plane waves. The same approximation is made on the other side. Wave components arriving at P from different parts of A are out of phase by an amount which varies linearly. This means that the terms arising from curvature of the wavefronts are being neglected.

Let us now perform the integration. For the square surface we considered above the x and y components can be independently integrated, and we find

$$\int_{-\Delta_y/2}^{\Delta_y/2} e^{-ik(Yy)} dy = \frac{-e^{-ik(Y\Delta_y/2)} + e^{ik(Y\Delta_y/2)}}{ikY} \quad (1.54)$$

$$= \frac{2 \sin(kY\Delta_y/2)}{kY} \quad (1.55)$$

$$= \Delta_y \frac{\sin(\beta_y)}{\beta_y} = -\Delta_y \text{sinc}(\beta_y) \quad (1.56)$$

where $\beta_y = kY\Delta_y/2$. Performing both integrals, and defining $\beta_x = kX\Delta_x/2$ we find

$$U(P) = K e^{ik(r_0+s_0)} \Delta_x \Delta_y \frac{\sin(\beta_y)}{\beta_y} \frac{\sin(\beta_x)}{\beta_x} \quad (1.57)$$

and the intensity (which is what we measure) is given by

$$|U(P)|^2 \propto I(P) = I(0) \frac{\sin^2 \beta_y}{\beta_y^2} \frac{\sin^2 \beta_x}{\beta_x^2} \quad (1.58)$$

where $I(0)$ is the maximum of the intensity. This function is plotted for one dimension, ie. with $\beta_y = 0$ in figure 1.9 for three different values of Δ_x . What we see is that as Δ_x increases with respect to λ , the function narrows. In the limit that $\lambda/\Delta_x \rightarrow 0$, we would obtain a Dirac delta function. This is the limit given by *geometrical optics*, where we can draw lines instead of worrying about waves. We'll use this to derive Snell's law below.

Following the previous section, we can modify the situation we considered earlier by placing lenses between the source and the aperture, and between the aperture and the destination point. This is shown in figure 1.10. If the source and destination are at focal points of the lenses, then the waves after the first lens (approaching the aperture) will be planar waves. After the aperture, any angular deviations are focussed down by the second lens to a single point on the image plane. The Fraunhofer approximation we made in equation 1.51, that the wavefront phase varies linearly across the aperture is then an *exact* condition.

Alternative statement of Fraunhofer condition

Above we stated that we would observe Fraunhofer diffraction if the phase of the incoming wave varies linearly across the aperture, and if we are able to separate angular components after the aperture. A more general condition is the following:

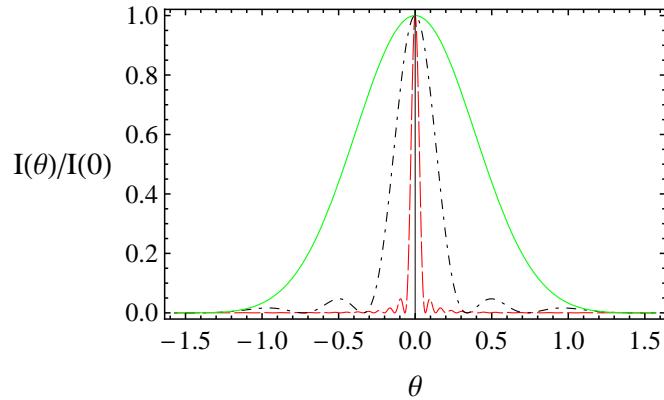


Figure 1.9: Fraunhofer diffraction patterns $I(\theta)/I(0)$ for $\Delta_x/\lambda = 15$: red, dashed, $\Delta_x/\lambda = 3$: black, dot-dashed and $\Delta_x/\lambda = 1$: green, solid. Note that the peak narrows dramatically as the slit width increases.

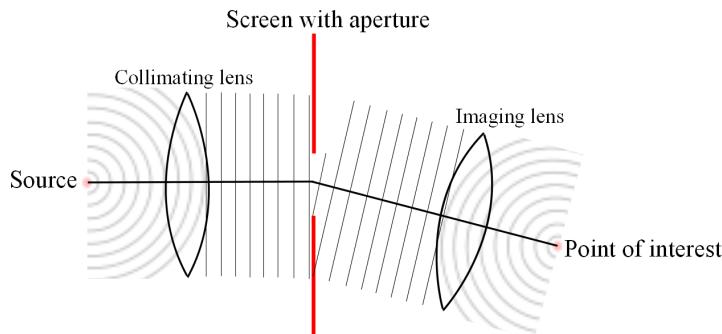


Figure 1.10: Use of a lens to ensure that the Fraunhofer condition is met with no approximations. Due to the collimating lens, plane waves are now incident on the aperture. The waves which arrive at a particular point in the imaging plane of the second lens must all be propagating in the same direction before this lens - they are planar wavefronts.

- Fraunhofer diffraction is observed in the image plane of the object. Here the diffraction is caused by apertures which appear between the source and the image plane.

This definition makes clear that Fraunhofer diffraction will be especially important for imaging instruments. Any place where we image a source, we should be aware that any sort of aperture restricting the path of light between the source and the image will cause Fraunhofer diffraction. This is one of the primary limits to resolution in optical instruments.

The above definition may not seem to be 100 % clear from the physics we described earlier (to me, anyway). What if the aperture occurs at a point where the wavefronts are very curved, couldn't we then violate our approximations? The answer is no, and is best illustrated by figure 1.11. Here our thought experiment involves two (theoretical) lenses with opposite focal lengths placed immediately before and immediately after the aperture. The first makes the incoming wavelengths planar, while the second puts them back on their original path. Clearly as these lenses approach each other then in the thin-lens approximation they are doing nothing, except that the aperture is working on planar wavefronts.

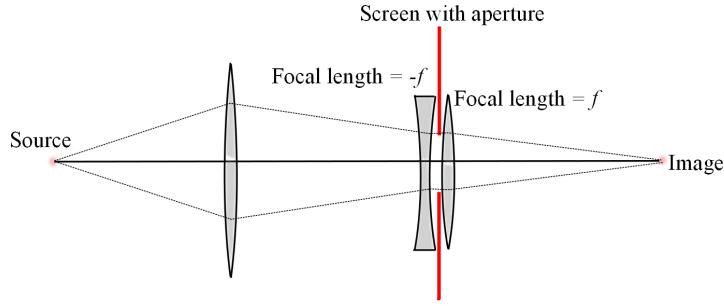


Figure 1.11: Thought experiment for second statement of the Fraunhofer condition. Fraunhofer diffraction is observed in the image plane of the source.

Fraunhofer diffraction and Fourier Transforms

The second situation we might consider is that instead of A being an aperture, it is a pattern of apertures. We can define such a pattern by $f(x, y)$, where the function is 1 if the point x, y transmits perfectly, and 0 if it is opaque. The use of the function also allows us to introduce regions of partial transmission, or a phase shift (for this, f must be complex). Instead of equation 1.51, we now get

$$U(P) = K e^{ik(r_0+s_0)} \int_{-\Delta_x/2}^{\Delta_x/2} \int_{-\Delta_y/2}^{\Delta_y/2} f(x, y) e^{-ik(Xx+Yy)} dx dy \quad (1.59)$$

which we see is of the same form of the Fourier transform which you met last term. It is a two-dimensional transform, with the integral taken over the area A . Thus in Fraunhofer diffraction, the resulting intensity pattern after focussing by a lens on a screen is the Fourier transform of the aperture. This is one area (there are many) where Fourier transforms are extremely useful. You will use these methods in the problem sets for more complex problems.

1.2.5 Huygen's principle and refraction

The other important result you came across already (and which plays a key role in imaging) is refraction. One typically derives Snell's law by looking at the interface of two different materials, with plane waves propagating in each material, or by the use of Huygen's principle, where you imagine a row of sources on the interface which emit spherical waves at the same phase as they are excited by the incident light. Actually, Huygen's principle is described very well by the Kirchoff integral. The s components of the integral are the incident field, so they tell the Huygen's sources what phase to emit. The r component gives the contribution to the field at P from these spherical sources. As we integrate over all these sources, we do exactly what Huygen told us to do.

To illustrate this, let us use our integral formalism to derive Snell's law for refraction. Imagine that instead of the surface A separating two regions of free space, the inside of our original closed surface (enclosed by A , B and C) is filled with a material with a different refractive index, n_1 . Then the wavevector in this region will be modified to $k_1 = n_1 k$. The simplified form of the Kirchoff-Fresnel integral then becomes

$$U(P) = \frac{iA}{4\pi r_0 s_0} (k_1 \hat{r}_0 \cdot \hat{n} - k \hat{s}_0 \cdot \hat{n}) \int_A e^{iks + ik_1 r} da = K \int_A e^{iks + ik_1 r} da \quad (1.60)$$

Now consider that we take the Fraunhofer approach, but over a region A which has dimensions large compared to the wavelength. The intensity distribution will now be given by an

integral similar to that in equation 1.51, only we have to take care of k and k_1

$$U(P) = K e^{i(k_1 r_0 + k s_0)} \int_{-\Delta_x/2}^{\Delta_x/2} \int_{-\Delta_y/2}^{\Delta_y/2} e^{-i(X' x + Y' y)} dx dy \quad (1.61)$$

where

$$X' = k \frac{x_S}{s_0} + k_1 \frac{x_P}{r_0} \quad (1.62)$$

$$Y' = k \frac{y_S}{s_0} + k_1 \frac{y_P}{r_0} \quad (1.63)$$

and we should note here that k and k_1 appear in these conditions where they did not in equations 1.52 and 1.53. Performing the integral gives us a sinc function centred at $X' = 0$, $Y' = 0$, as we found previously. This provides a condition in the x component which looks like

$$k \frac{x_S}{s_0} = -k_1 \frac{x_P}{r_0} \quad (1.64)$$

$$\frac{n\omega}{c} \sin(\theta_S) = \frac{n_1 \omega}{c} \sin(\theta_P) \quad (1.65)$$

$$n \sin(\theta_S) = n_1 \sin(\theta_P) \quad (1.66)$$

which is Snell's law. Note that this only gives the angle at which the centre of the sinc function appears. However, we should also note that we integrated over a region with dimensions large compared to the wavelength. This means that our sinc function is a good approximation to a delta function - all the light propagates at approximately the refraction angle.

1.3 Spatial Resolution

1.3.1 Telescopes and the Rayleigh criterion

What is the resolving power of a telescope? The limit is due to diffraction. Consider two light sources coming through an aperture (the diameter of your optical system is one such). Of course the light from each of these objects will diffract, and produce a pattern similar to that of equation 1.58. This intensity pattern will appear on a CCD camera, or on our eye. If the two objects are very close to each other, the two intensity patterns will merge, and we cannot resolve the objects. What we can resolve depends on how sensitive we are to drops in intensity. The most famous criterion for resolution is due to Rayleigh, and states that

- Two objects are just resolved if the maximum from the intensity pattern due to one falls on the first minimum of the intensity pattern due to the other

Let us consider the case that all the diffraction in a telescope is due to the limited aperture formed by the diameter D of the front lens of the telescope, then the first minima in the diffraction pattern is at an angle where $\sin(kXD/2) = \sin(kD \sin \theta/2) = 0$, giving

$$\theta = \frac{\lambda}{D} \quad (1.67)$$

This gives the minimum difference in angles between sources that could be resolved by the telescope. You note that making D big is of great advantage for telescopes, which is why astronomical telescopes are so large. It is also worth noting that if you want to look at low frequency sources, the wavelengths become longer, and it becomes harder to resolve

the position of the source. A final point. We used the Rayleigh criterion above, which is somewhat basic, but other measures do not produce markedly different results to other criteria which you may find in the literature (the amount of intensity drop which is required to resolve separate peaks depends on many details of the measurement apparatus which is being used, so there is little point in us trying to be more rigorous here).

We did make one assumption above which I didn't mention: I assumed that the light from the two different sources does not interfere. This allowed us to place the intensity patterns on top of each other and not worry about the relative phase of the oscillating fields arriving at the detector. This is quite a good approximation for two stars, since there is no reason that the light should have similar polarization, stable relative phase etc.. However if our method of observing objects is to shine light from a single source at them, we should worry about this possibility. We will look at an example of this in the next section.

1.3.2 Microscopy and the Abbé criterion

Diffraction plays an especially important role in limiting the performance of the microscope. In a microscope, the object under study is illuminated by a single source of light. Thus we should not ignore coherence (a phase relationship) between light coming from different parts of the object.

In order to illustrate the method by which the limit to resolution of a microscope can be found, consider that we have an opaque screen with an aperture, where the aperture is now filled by a structure with transmission given by $\sin^2(k_A x)$. This is an example object which we're interested in looking at with a well defined spatial frequency k_A . We now look at the image of this object, using an objective lens with diameter d_L and focal length f_L placed at a distance f_L from the aperture. In the far field limit, the amplitude of the field is given by the Fourier transform of the transmission function, because we are in the limit where Fraunhofer diffraction is appropriate.

$$A = \frac{K}{2} \int_{-a}^a e^{ikx \sin \theta} (1 + \cos(2k_A x)) dx \quad (1.68)$$

$$\propto \frac{\sin(ka \sin \theta)}{ka \sin \theta} + \frac{\sin(ka \sin \theta - k_A a)}{ka \sin \theta - k_A a} + \frac{\sin(ka \sin \theta + k_A a)}{ka \sin \theta + k_A a} \quad (1.69)$$

Looking at this form, we see that information about the periodicity of the partially transmitting structure is contained in the last two terms, which are sinc functions centered at $ka \sin \theta = \pm k_A a$. If we use a lens which doesn't span a wide enough angle to catch this light, we will lose all information about the period k_A . This gives us a criterion on the numerical aperture (NA) of the lens, which is the sine of the angle subtended by the lens at the object. In order to resolve the period k_A , we require $NA > \sin \theta_c = k_A/k$. Turning this around to a measure of resolution, we find that in this case the limiting distance between two resolved objects must be $\lambda/(NA)$. This is the Abbé limit of microscopy.

1.4 Polarization

Up until now, we've considered light as a scalar wave, but have primarily conveniently forgotten about the fact that light is the propagation of electric and magnetic waves, which are perpendicular to one another, and to the direction of travel. When a wave is traveling through free space, this is fine, because there's nothing to compare these directions to. However, if light is incident on a surface at a given angle, then suddenly the orientation of the surface will define a specific direction, and we must start to worry whether the direction of the electric and magnetic field vectors, otherwise known as the *polarization* of the light will matter. Note that once we know the direction of the electric field vector and the propagation vector, the magnetic field direction is fully specified.

1.4.1 Types of polarization

As light propagates, the electric field vector of the light can evolve in a number of different ways. To illustrate this, let us use an axis system where the light propagates in the z direction. As we know from electromagnetism, the electric field vector will be perpendicular to the propagation direction. Therefore the electric field can point in any direction in the $x - y$ plane.

Linear polarization

Linear polarization is when the field vector always points in the same direction. We can decompose the electric field in terms of components which lie along the x and y directions, ie.

$$\mathbf{E} = (E \cos(\theta) \hat{x} + E \sin(\theta) \hat{y}) \cos(kz - \omega t) \quad (1.70)$$

$$= (E_x \hat{x} + E_y \hat{y}) \cos(kz - \omega t) \quad (1.71)$$

where θ is the angle between the electric field vector and the x axis. In laboratory situations, it is common to align x and y with horizontal and vertical, and we refer to horizontal and vertical linear polarization.

Circular and elliptical polarization

An additional form of light occurs when there is a phase difference between the x and y polarizations. We will see how this might arise physically in the next section. The mathematical form of this type of polarization is

$$\mathbf{E} = E_x \hat{x} \cos(kz - \omega t) + E_y \hat{y} \cos(kz - \omega t + \alpha) \quad (1.72)$$

The electric field vector therefore points along different directions as time evolves. In addition, the magnitude of the field vector can change in time - the extreme case of this is when it periodically reaches zero, which is linear polarization. The time average of the electric field vector trajectory is shown for a few different values of E_y , E_x and α in figure 1.12. It can be seen that the principal axes of the polarization of the ellipse are not necessarily along x or y . One special case is worth mentioning. When $E_x = E_y$ and $\alpha = \pm\pi/2$ then we get circularly polarized light. For $\alpha = \pi/2$ we call this σ^+ polarization, whereas for $\alpha = -\pi/2$ we call it σ^- polarization. We will see these two forms later on, because circular polarization interacts with atomic transitions in a simple manner.

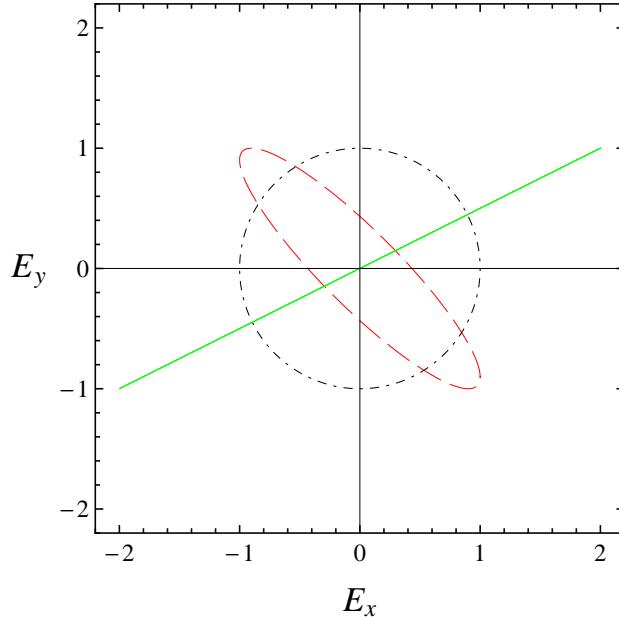


Figure 1.12: Time average of polarization for three different parameters sets of equation 1.72. The parameters are $(E_x, E_y, \alpha) = (1, 1, 3\pi/7)$: dashed, red, $(1, 1, \pi/4)$: dot-dashed, black, $(1, 2, 0)$: solid, green.

1.4.2 Birefringence

There are many materials which are not isotropic with respect to optical interactions. This is typically because the way in which the atoms are distributed in the material is not isotropic (the layers of atoms have different separations in different directions). For the purposes of our discussion at this particular moment, we shall not describe why, but I merely ask you to accept that this type of material exists. One result is that the speed of light in the material changes as a function of the polarization of the wave.

One common class of materials which are anisotropic in a simple manner are birefringent materials. These are characterized by two axes, perpendicular to one another, for which the refractive index of the light is different. Thus a light wave polarized along the fast axis travels faster than one polarized along the slow axis. If we make the x axis have refractive index n_x and the y axis have refractive index n_y , then on inputting a general polarized light field we find that after time t , and at distance z into the material, the electric field evolves to

$$\mathbf{E}(z, t) = E_x \hat{\mathbf{x}} \cos(k_x z - \omega t) + E_y \hat{\mathbf{y}} \cos(k_y z - \omega t + \alpha) \quad (1.73)$$

where $k_x = n_x k_0$ and $k_y = n_y k_0$. Since the phase difference of the two components will now evolve, the light polarization evolves in time. If we make a plate where the thickness of the material L is such that $(n_x - n_y)k_0 L = \pi$, then we call the plate a *half-wave plate*. If the input polarization is linear, a half-wave plate will rotate the axis of the polarization, but it will remain linear. The rotation angle is a function of the angle between the fast axis of the crystal and the input polarization.

For $(n_x - n_y)k_0 L = \pi/2$ we call the plate a *quarter-wave plate*. This is able to convert linearly polarized light into elliptically polarized light. This also depends on the angle between the polarization of the input light and the fast axis of the crystal. For example, if the input light is polarized along the fast axis, then there will be no change in the polarization. At 45 degrees to the fast axis, however, the quarter-wave plate will convert linear polarization into circular polarization.

1.4.3 Fresnel equations

How can we produce polarized light? Well, light from atoms is in many cases polarized, but only with respect to the environment of the atom which produces it. Since many atoms together very often have random polarization (such as atoms in a gas lamp), it is not common to get a polarized light source from a collection of atoms. In the world around us we typically find polarized light when the light reflects from a boundary between two media with different refractive index. You are probably most familiar with this when you buy polarized sunglasses, which are primarily of use at the seaside, because the sea has a very different refractive index to the air. In the next section, we will find out why this occurs, by looking at what Maxwell's equations have to tell us about the transmission of light through a surface.

We will now look at one example of the effects of polarization. For simplicity, we will restrict our discussion to plane waves, and look at a couple of simple cases where a plane wave is incident on a surface. I will analyze this situation for one polarization of light, the other you will cover in the problem sets.

Electromagnetic waves in dielectrics

Let's remind ourselves of the relation between the electric and magnetic fields in electromagnetic waves, but now in the presence of dielectric media. We start from Maxwell's equations in the absence of free charges and currents

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \quad (1.74)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (1.75)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (1.76)$$

$$\nabla \cdot \mathbf{D} = 0 \quad (1.77)$$

Taking the curl of both sides of equation 1.75, substituting $\mathbf{B} = \mu_r \mu_0 \mathbf{H}$, and using equation 1.74 we quickly find

$$-\nabla^2 \mathbf{E} = -\mu_r \mu_0 \epsilon_r \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (1.78)$$

which is of the form of a wave equation with wave propagation speed given by

$$v = \frac{1}{\sqrt{\mu_r \mu_0 \epsilon_r \epsilon_0}} = \frac{c}{n} \quad (1.79)$$

where c is the speed of light in vacuum and $n \equiv \sqrt{\mu_r \epsilon_r}$ is the refractive index of the material.

One solution to the wave equation is a field

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 (e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} + e^{-i(\mathbf{k} \cdot \mathbf{r} - \omega t)}) \quad (1.80)$$

The magnetic field component which propagates along with this electric field can be found from equation 1.75. Taking the curl of $\mathbf{E}(\mathbf{r}, t)$ and then integrating to get \mathbf{B} we find

$$\mathbf{B}(\mathbf{r}, t) = \frac{\mathbf{k} \times \mathbf{E}(\mathbf{r}, t)}{\omega} \quad (1.81)$$

With these solutions in hand, we have the tools to investigate the propagation of electromagnetic waves between different types of dielectric media.

Boundary conditions

The first component which is required to understand the propagation of polarized light regards the boundary conditions for electric and magnetic fields at the interface between two materials.

Let us start by considering the rules for magnetic fields. We can use Gauss' law to write

$$\int_V \nabla \cdot \mathbf{B} d^3r = \oint_S \mathbf{B} \cdot \mathbf{n} da \quad (1.82)$$

Let us consider the field component perpendicular to a surface, and choose an enclosed volume where the top half lies in material 1, and the bottom in material 2 (see figure 1.13 a)). Ignoring the sides of the volume, the top and bottom surfaces give $\mathbf{B}_1 \cdot \mathbf{n} = B_1^\perp$ and $\mathbf{B}_2 \cdot \mathbf{n} = -B_2^\perp$. In order to ignore the sides, we let the thickness of the volume tend to zero, which also takes it's volume to zero. Therefore we find from the Maxwell relation $\nabla \cdot \mathbf{B} = 0$ that

$$B_1^\perp = B_2^\perp \quad (1.83)$$

i.e. the component of the magnetic field perpendicular to the surface is continuous across the surface.

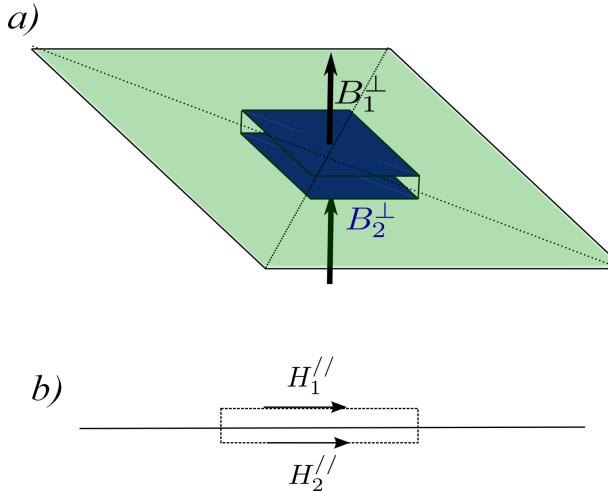


Figure 1.13: Boundary conditions for electric and magnetic fields at the interface between dielectric media. a) The perpendicular component of \mathbf{B} . For an infinitesimally thin volume, all the magnetic field must come out through the top and bottom. The components normal to the surface must therefore equal one another b) Parallel component of \mathbf{H} . If the sides of the loop tend to zero, the only contributions come from the two parts of the loop parallel to the boundary.

The second relation we have for the magnetic fields comes from the other Maxwell equation

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \quad (1.84)$$

Here we will integrate over a surface, so will make use of the Stokes theorem for the left hand side

$$\int_A (\nabla \times \mathbf{H}) \cdot \mathbf{n} da = \oint_S \mathbf{H} \cdot d\mathbf{l} \quad (1.85)$$

Now, instead of the geometry chosen in the previous paragraph, we choose the loop shown in figure 1.13 b), which passes for a distance L parallel to the surface on one side, then crosses

the surface, then travels back parallel to the surface in the other material. It is closed by again crossing the surface. We note that for the parts of the path parallel to the surface we get $\mathbf{H}_1 \cdot d\mathbf{l} = H_1^{\parallel} dl$, $\mathbf{H}_2 \cdot d\mathbf{l} = -H_2^{\parallel} dl$. Now letting the edges of the loop perpendicular to the surface tend to zero length, and noticing that the area of the loop and thus the integral of $\partial\mathbf{D}/\partial t$ over the area also goes to zero as long as $\partial\mathbf{D}/\partial t$ is finite, we find that

$$H_1^{\parallel} = H_2^{\parallel} \quad (1.86)$$

You can easily show by similar methods (go do it!) that the relations which hold for the electric components are

$$E_1^{\parallel} = E_2^{\parallel} \quad (1.87)$$

$$D_1^{\perp} = D_2^{\perp} \quad (1.88)$$

Now we will use these relations to look at the reflection and transmission co-efficients of plane waves impinging on a surface. The physics of this is similar to what you saw if you looked at transmission lines in your course last term. Essentially the change in refractive index requires us to meet a phase matching condition for the light. We consider light as plane waves incident at an angle θ_i to a boundary between two materials. Both materials are linear, in the sense that $\mathbf{D} = \epsilon_r \epsilon_0 \mathbf{E}$ and $\mathbf{B} = \mu_r \mu_0 \mathbf{H}$. The material in which the light starts is has $\mu_r = \mu_1, \epsilon_r = \epsilon_1$ and the second material $\mu_r = \mu_2, \epsilon_r = \epsilon_2$.

The other case, for which the electric field vector points out of the plane is left as a homework exercise, so that you learn about this properly (by doing, not watching). We consider three

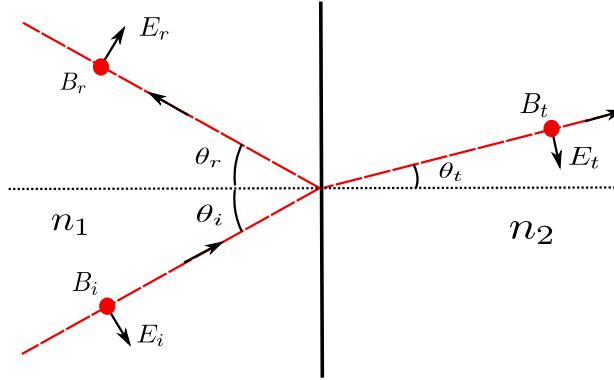


Figure 1.14: Incident, transmitted and reflected wave parameters for radiation at an interface between materials with different refractive indices. The electric field is polarized perpendicular to the plane of incidence, and thus points out of the page. The magnetic field components are perpendicular to both the electric field and propagation direction, as is usual for an electromagnetic wave.

waves shown in figure 1.14. The first is the incident wave, which propagates towards the surface with angle of incidence θ_i . At the surface, this generates two waves, one of which is reflected and stays within the first medium, and one which is transmitted into the second region. Taking the electric field components parallel the surface, we have

$$E_i \cos \theta_i - E_r \cos \theta_i = E_t \cos \theta_t \quad (1.89)$$

while for those perpendicular to the surface we have

$$\begin{aligned} \epsilon_1 (E_i \sin \theta_i + E_r \sin \theta_i) &= \epsilon_2 E_t \sin \theta_t \\ n_1^2 (E_i \sin \theta_i + E_r \sin \theta_i) &= n_2^2 E_t \sin \theta_t \end{aligned} \quad (1.90)$$

The magnetic components are parallel to the surface, and are therefore given by $B_i = n_1 E_i / c$, $B_r = n_1 E_r / c$, and $B_t = n_2 E_t / c$, which using the boundary condition 1.86 (while remembering that $\mu_1 = \mu_2$) gives

$$n_1(E_i + E_r) = n_2 E_t . \quad (1.91)$$

The first thing to note is that by combining this expression with equation 1.90 along with the reflection condition that $\theta_i = \theta_r$, we recover Snell's law. Well, we knew Snell's law already, and have gained no extra information about the relative sizes of the transmitted and reflected waves, which is what we desire. So let us instead combine the magnetic field condition with equation 1.89. We find that

$$\frac{E_r}{E_i} = \frac{n_2 \cos \theta_i - n_1 \cos \theta_t}{n_2 \cos \theta_i + n_1 \cos \theta_t} \quad (1.92)$$

$$\frac{E_t}{E_i} = \frac{2n_1 \cos \theta_i}{n_2 \cos \theta_i + n_1 \cos \theta_t} \quad (1.93)$$

These equations are very similar to those you would find if you do problems on impedance matching in transmission lines. If you think about this, we shouldn't be so surprised - a transmission line is just propagating an electromagnetic wave after all! What we are saying now though, is that impedance matching of materials is useful for getting light in and out of materials, and it depends on the polarization of the light.

Brewster's angle

Let us take a look at the result in equation 1.92 in a little more detail. It helps to do some trigonometric algebra, which leads us through

$$\begin{aligned} \frac{E_r}{E_i} &= \frac{n_2 \cos \theta_i - n_1 \cos \theta_t}{n_2 \cos \theta_i + n_1 \cos \theta_t} \\ &= \frac{\sin \theta_i \cos \theta_i - \sin \theta_t \cos \theta_t}{\sin \theta_i \cos \theta_i + \sin \theta_t \cos \theta_t} \\ &= \frac{\sin 2\theta_i - \sin 2\theta_t}{\sin 2\theta_i + \sin 2\theta_t} \\ &= \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \end{aligned} \quad (1.94)$$

Why did we go to all the trigonometry to get to this result? The reason is that it becomes obvious from this last expression that if $\tan(\theta_i + \theta_t) \rightarrow \infty$ then $E_r/E_i \rightarrow 0$, ie. no light will be reflected. The condition for this to occur is of course that $\theta_i + \theta_t = \pi/2$. Inserting this form into Snell's law, we can use the refractive indexes of the two materials to find the angle of incidence θ_{iB} for which no light is reflected

$$\tan \theta_{iB} = \frac{n_2}{n_1} \quad (1.95)$$

We call this critical angle Brewster's angle. It is important because if unpolarized light (made up of light with electric field vectors in random directions) is incident on a surface at this angle the reflected light will now be polarized, because only light which is polarized parallel to the interface is reflected.

It is worth recalling your trips to the seaside again (or the lake if you are Swiss). The light which is reflected will have which polarization? What does this mean about polarized sunglasses?

1.5 Spectral measurements of light

The discussion of the previous section exclusively concerns the propagation of *monochromatic* light through space. In many optical instruments, we use the differences in spatial propagation for different colours of light in order to deduce information about the spectral composition of the light. There are two effects which can give rise to differences in the spatial propagation of light. The first is dispersion; that in most materials light of different colours propagates at different speeds. This means that *refraction* can be used to look at the spectrum, and is the working principle of a prism. The second is in diffraction: note that λ appears in many of the expressions we see for diffraction and interference patterns. This means that the interference patterns for different colours are different, and comes about because the phase acquired over a propagation distance $|z|$ is $n|z|/\lambda$. Interference is one of the most powerful tools we have for looking at spectra.

The most general problem in spectra is that we have an unknown light source, and we want to know what set of colours it produces. You might say well this is easy, my eye does this, just use three different detectors which are only sensitive to different colours. But at some point you want to be more precise, and have better spectral resolution. It was the investigation of spectra which provided many of the experimental results from which quantum mechanics was developed. A particularly famous example is the spectrum of atomic hydrogen, which we will come to later in this course.

1.5.1 The Prism

The most familiar device which separates out different colours is the prism. The prism makes use of Snell's law and the fact that different colours have different speeds in materials. For light propagating from a vacuum into a medium with refractive index n , we have that $\theta_j = \arcsin(n \sin \theta_i)$, so if $n = n(\lambda)$ is a function of angle, θ_j will also be. Materials used for modern prism spectrometers are SiO, ZnSe and NaCl. Prism spectrometers have the advantage that you typically see the whole spectrum, but the ability to separate nearby colours is limited. This is because refractive indexes of materials do not vary a lot across the visible spectrum (hence nearby colours are not separated in angle by very much). For example, in Crown glass (one of the types which is produced), the refractive index varies by approximately 2% between blue and red light. As a result of the fact that they are not commonly used in modern physics, we will concentrate on other spectroscopic devices and not spend any more time on the prism.

1.5.2 The Grating Spectrometer

The grating spectrometer is a device which finds common use today in Chemistry and Physics (and possibly other sciences). It uses the principle of multiple slit interference to separate out spectral components. In many cases a reflective grating is used instead of letting light pass through a transmission grating, as in the multi-slit experiments you studied last term. In order to produce an object which periodically reflects light, it is common to scribe non-reflecting lines on a reflective metal surface. The physics of both the transmission and reflection gratings is much the same, as we shall see. The arrangement of the spectrometer is drawn in figure 1.15.

Light from an extended source is focussed onto a slit. This slit appears as an apparent "point" source of light, and is what will be imaged on the detector. After the slit, the light is collimated by a lens one focal length away such that planar wavefronts fall upon the reflective grating at an angle of incidence θ_i . The light is reflected from the grating, and all

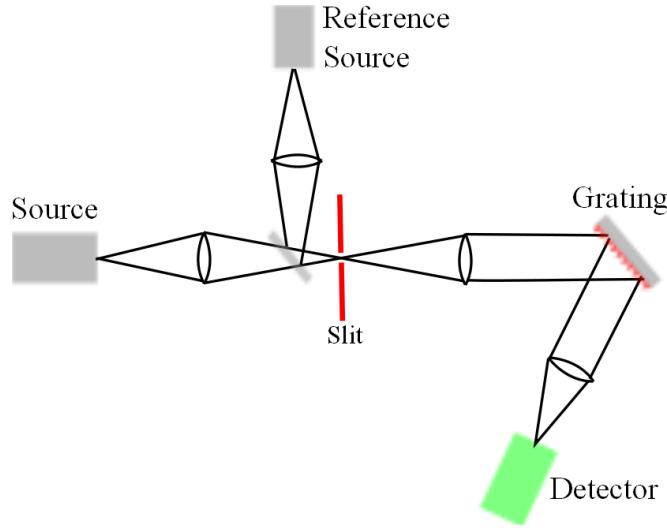


Figure 1.15: The grating spectrometer. The source light is focussed onto an entrance slit, and then collimated by a lens placed after the slit. The now collimated light is reflected from the grating, and subsequently focussed down onto the detector (typically an extended detector such as a CCD camera). The reference source provides information about the absolute frequency of the light, because its spectrum will overlay the unknown source spectrum.

light which travels at angle θ_j is focused down by a second lens onto a single point. By placing a movable slit in the focal plane of this second lens, we can observe the change in intensity as a function of the position (otherwise we can use a CCD camera). The distance between adjacent reflective segments is d , and thus the phase difference between light incident at θ_i and reflected at an angle θ_j coming from two adjacent reflective strips on the grating can be calculated as

$$\delta\phi = kd(\sin(\theta_j) - \sin(\theta_i)) . \quad (1.96)$$

(figure 1.16 should be useful in helping you to see this). If the grating consists of N lines,

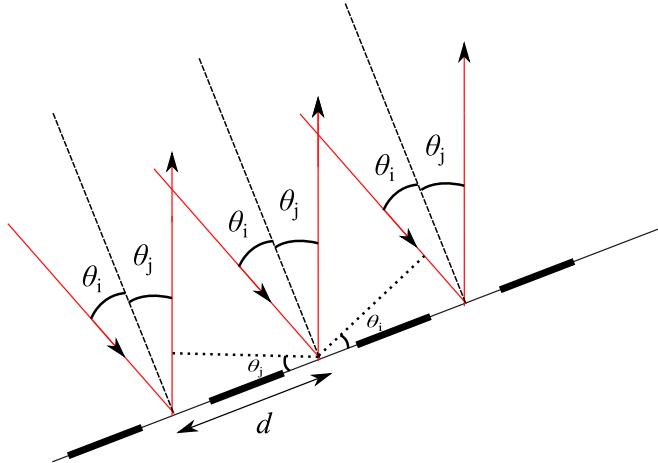


Figure 1.16: Optical paths in reflection from a grating with reflective segments spaced by distance d . The light incident at angle θ_i is reflected, but due to interference will not necessarily be reflected at θ_i . The path difference between light reflected from neighboring segments at angle θ_j can be calculated from the quantities in this figure, and is given in the main text.

then instead of integrating over a continuous surface as we did in our studies of diffraction,

we make a sum over amplitudes

$$A(\theta) = \sum_{n=-N/2}^{N/2} e^{in\delta\phi} \quad (1.97)$$

which, similar to the case of multiple-slit interference, gives an intensity distribution

$$|A(\theta)|^2 = |A(0)|^2 \frac{\sin^2(\delta\phi N/2)}{\sin^2(\delta\phi/2)} \quad (1.98)$$

Now a reminder of the principal features of this form. At the center of the pattern we get a maximum when $\theta_j = \theta_i$: this is just standard reflection, and gives no spectral information. However we will also get principle maxima for $\delta\phi/2 = p\pi$, since the denominator will go to zero. The adjacent zeros of the function occur when $N\delta\phi/2 = (pN + 1)\pi$ (i.e. when the numerator next goes to zero).

We wanted to use this as a spectral device, ie. we want to see whether it can separate out different values of λ . Since we're trying to separate these two components in space, this is a resolution question, but now we see it posed in terms of spectral resolution. What if we apply a similar criterion here as we did for spatial resolution, but turn it around? I can then resolve two colours if the peak from the pattern of wavelength 1 falls at the minimum adjacent to the peak from the pattern of wavelength 2.

At this point we have two conditions

$$\frac{\delta\phi(\lambda_1)}{2} = p\pi \quad (1.99)$$

$$\frac{\delta\phi(\lambda_2)N}{2} = (Np + 1)\pi \quad (1.100)$$

We can now subtract these two expressions, and find

$$Nd(\sin(\theta_j) - \sin(\theta_i)) \left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right) = 1 \quad (1.101)$$

which making the definition $\delta\lambda = \lambda_1 - \lambda_2$ and the assumption $\lambda \simeq \lambda_1, \lambda_2$ gives

$$\frac{\delta\lambda}{\lambda} = \frac{\lambda}{Nd(\sin(\theta_j) - \sin(\theta_i))} \quad (1.102)$$

As we pursue science at higher precision, we wish to ask “how far can I push this device?”. In the case of a spectrometer, the relevant question is “what determines the closest wavelengths which I can separate out in my apparatus?”. We can easily see from trigonometry that the maximum value of $(\sin(\theta_j) - \sin(\theta_i))$ that is possible is 2, so the maximum possible resolution with this device is

$$\frac{\delta\lambda}{\lambda} = \frac{\lambda}{2Nd} \quad (1.103)$$

We see one of the most important results that you should remember for the rest of your days. The closest spectral components which can be resolved in a spectrometer are given by the reciprocal of the maximum path difference between light passing through different parts of the apparatus. In the case of the grating, this is the complete distance across the grating, which is given by Nd . We will see this result again in the context both of the Michelson interferometer and the Fabry-Perot etalon, which we will meet in the next sections. It is true for any spectroscopic instrument.

In many uses of grating spectrometers, they are setup so that one particular order of the grating is used, or phrased in a more experimental way, only a small range of angles on

either side of a given angle satisfying equation 1.99 are used. In this case, equation 1.102 tells us that the resolution will be

$$\frac{\delta\lambda}{\lambda} = \frac{1}{Np} \quad (1.104)$$

The inverse of this quantity is often known as the *resolving power* of a grating, since it goes up (the grating is more powerful) as the ability to resolve closely spaced spectral lines increases.

1.5.3 The Michelson Interferometer

The Michelson interferometer is one of the classic measurement devices of optics. It was originally devised as a way of measuring the direction dependence of the speed of light, as a test of the direction with which the Earth was traveling through a mystery background called the “aether”. In fact, it is still being used today to test General Relativity, at the LIGO (Laser Interferometer Gravitational-wave Observatory) experiment. The layout of a

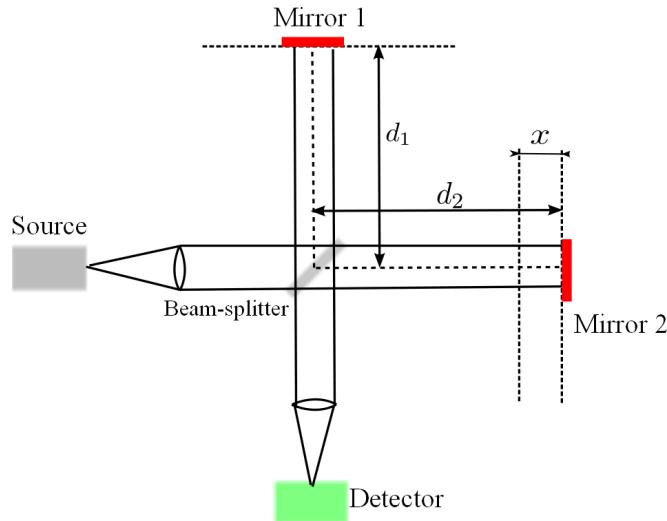


Figure 1.17: The Michelson interferometer. Light from a source is split into two paths, each of which is reflected from a mirror. The light returning from the two mirrors is recombined on the beam splitter, and then focussed down onto the detector. When operated as a spectrometer, the distance $x = d_2 - d_1$ is scanned and the intensity of the light at the detector is recorded as a function of x .

Michelson interferometer is shown in figure 1.17. It works by splitting light equally between two different paths using a beam-splitter. At the opposite end of each arm of the Michelson lies a mirror, which retro-reflects the light so that it arrives back at the beam-splitter. The two light beams then interfere. If the beam splitter-mirror distances of the two arms are d_1 and d_2 , then the phase difference between light which has taken the two paths is

$$\delta\phi = 2k(d_2 - d_1) \quad (1.105)$$

where k is the wavelength of the light. The amplitude of the light after light from the two paths are recombined on the beam splitter is

$$A = \frac{A_0}{2} (1 + e^{i\delta\phi}) \quad (1.106)$$

which gives an intensity distribution

$$\frac{|A|^2}{|A_0|^2} = \frac{1}{2} (1 + \cos(\delta\phi)) \quad (1.107)$$

Michelson “Fourier transform” spectrometer

If the source of light is made up of many colours and we open our detector for times long compared to the difference in these colours (and thus interference of light of different colours can be neglected) then if the power between frequencies f and $f + \delta f$ is given by $S(f)df$ the intensity distribution will be

$$I(x) \propto \int df \frac{S(f)}{2} (1 + \cos(4\pi fx/c)) \quad (1.108)$$

$$\propto \frac{I_0}{2} + \int df S(f) \cos(4\pi fx/c) . \quad (1.109)$$

The first term is an integral of the total power in the source, gives a constant which is independent of x . The second term, however, is proportional to the cosine Fourier Transform of the spectral density. Since the interference trace is symmetric, we can invert the Fourier transform to find

$$S(f) \propto \int_0^{x_{\max}} (2I(x) - I_0) \cos(4\pi fx/c) dx \quad (1.110)$$

where x_{\max} is the maximum path difference between the two arms of the interferometer. This expression tells us that the spectral density can be obtained by Fourier transforming the recorded intensity pattern as a function of x .

This seems great, but now you ask, how good is the resolution? Let's recall our arguments from Fraunhofer diffraction about spatial resolution, and see if we can turn them into arguments regarding spectral resolution. In Fraunhofer diffraction, when the slit increased in size the peak would get narrower. Now we are asking a slightly different question, but with a very similar integral: a monochromatic source would produce an intensity pattern which is given by equation 1.107 as a function of x . If the monochromatic source has center frequency f_0 , and we use equation 1.110 to perform the Fourier transform, we will get

$$S(f) \propto \int_0^{x_{\max}} \cos(4\pi f_0 x/c) \cos(4\pi fx/c) dx \quad (1.111)$$

$$\propto \frac{\sin(4\pi(f - f_0)x_{\max}/2)}{4\pi(f - f_0)/c} + \frac{\sin(4\pi(f + f_0)x_{\max}/2)}{4\pi(f + f_0)/c} \quad (1.112)$$

Since we are dealing with optical frequencies which are closely spaced around a particular value, $f - f_0 \ll f + f_0$ and the second term will typically be negligible. The distance between the maximum and the nearest zero will be $\delta f = c/(4x_{\max})$, which if we make our criteria for resolution the same as that for the Rayleigh limit of spatial resolution means that frequencies differing by more than this amount will be resolved. Note that the largest path length between the two paths is $2x_{\max}$. This should ring bells - similar to the grating spectrometer, the resolvable frequency difference is given by c divided by the maximum path length achievable in the apparatus. This gives the Michelson a significant advantage over gratings, since it is easier to scan a mirror over a long distance than it is to build a very wide grating. This should give you a stronger hint now that you can immediately look at any spectral apparatus and estimate its spectral resolution. You just need to figure out the greatest path difference between light which is achievable in the apparatus.

1.5.4 The Fabry-Perot Etalon

The Fabry-Perot Etalon is perhaps the most widely used type of optical instrument around today. This is because it is not just used for spectroscopy, but is also one of the key components of the laser. The idea is simple. By using two highly reflecting mirrors which face each other, we can achieve a long differential path length (which we have seen is important

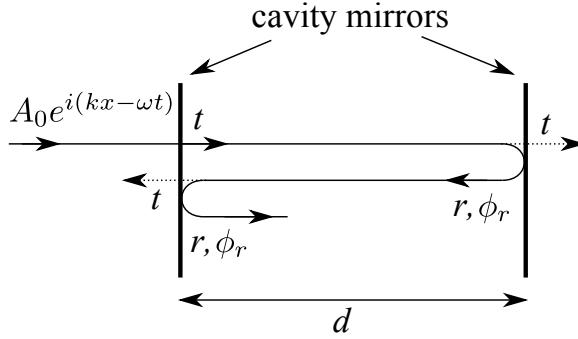


Figure 1.18: Schematic of a Fabry-Perot with mirror amplitude transmission co-efficient t and reflection co-efficient given by $re^{i\phi_r}$.

for high resolution in the Michelson and grating spectrometers) by passing light through the same region many times.

Let us imagine light incident from the left onto an etalon built by two identical parallel mirrors (I will sometimes refer to it as an optical cavity, which is the more common name today). A schematic is shown in figure 1.18. The light comes in at zero angle. For an initial field amplitude A_0 , the amplitude with which light is transmitted into the cavity is reduced by a factor t , therefore at $z = 0$ the amplitude due to the initial field is $A_0 t$ (note that t may be complex - ie. the light is phase shifted, but we shall ignore this here because it doesn't affect the results). Light propagates over the length of the cavity d , and is reflected by the mirror at the far end, changing the amplitude by a factor $re^{i\phi_r}$, where r is real and ϕ_r is the phase change on reflection. The light then travels back down the length of the cavity, and is reflected by the first mirror. This light is now travelling in parallel with the initial wave, and will add to the amplitude an amount $A_0 t r^2 e^{i\delta\phi}$, where

$$\delta\phi = k2d + 2\phi_r \quad (1.113)$$

Following each additional round trip we gain an additional contribution to the amplitude, such that the total amplitude looks like

$$A = A_0 t \sum_{n=0}^{\infty} r^{2n} e^{in\delta\phi} \quad (1.114)$$

Summing over infinite series is something we learnt long ago, so we can perform the sum, and find

$$A = A_0 t \frac{1}{1 - r^2 e^{i\delta\phi}} \quad (1.115)$$

In spectroscopy we typically look at the intensity of light emitted at the right hand side of the cavity. This is the same amplitude as found above, but with an additional factor te^{ikd} to account for the final trip across the cavity and the transmission of the right-hand mirror. Writing the emitted amplitude as $A_e = te^{ikd} A$, we find that the intensity of the outgoing light relative to that which was incident on the cavity is given by the expression

$$\begin{aligned} \frac{|A_e|^2}{|A_0|^2} &= \frac{t^4}{(1 - r^2 e^{i\delta\phi})(1 - r^2 e^{-i\delta\phi})} \\ &= \frac{t^4}{(1 + r^4 - 2r^2 \cos(\delta\phi))} \end{aligned} \quad (1.116)$$

To simplify this a little, let us make the assumption that the light is either transmitted or reflected by the mirrors (no light is absorbed). This is very often a pretty good assumption.

Then we have a relation $t^2 = 1 - r^2$. Let us also move to using a different set of constants describing the fraction of intensity which is reflected or transmitted, and call these $R \equiv r^2$ and $T = (1 - R) \equiv t^2$. Then we can write

$$\begin{aligned} \frac{|A_e|^2}{|A_0|^2} &= \frac{(1-R)^2}{(1-R)^2 + 2R(1-\cos(\delta\phi))} \\ &= \frac{1}{1 + \frac{4R}{(1-R)^2} \sin^2(\delta\phi/2)} \end{aligned} \quad (1.117)$$

This may all have seemed a bit arbitrary, which it was, because we were chasing in particular

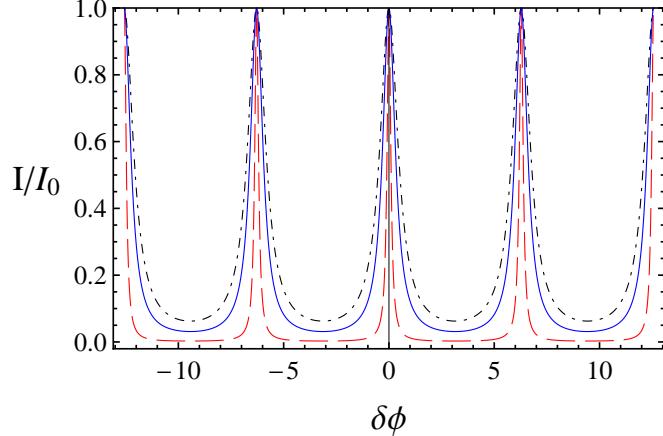


Figure 1.19: Fringe pattern as a function of phase difference for a Fabry-Perot etalon. Peaks occur for $\delta\phi = 2\pi m$ where m is an integer. The mirror reflection is $R = 0.9$: red, dashed, $R = 0.7$: blue, solid, $R = 0.6$: black, dot-dashed. There are many ways in which $\delta\phi$ can be varied in a Fabry Perot. Initial work looked at the angular distribution of light coming from the etalon. Pressure scanning of the gas between the plates is another method.

this result. What is the significance of this result? Well let us plot the transmitted intensity as a function of $\delta\phi$ for a few different values of R . The results are shown in figure 1.19. We see a set of peaks which are regularly spaced by $\delta\phi = 2\pi m$ where m is an integer, and have a width which decreases rapidly with R . The width and spacing are two quantities which are important when making measurements with a Fabry-Perot. They tend to be quoted as spectral quantities, ie. phrased in the language of frequency. The distance between adjacent peaks in frequency is found by noting that

$$\delta\phi_{\nu_1} - \delta\phi_{\nu_2} = 2\pi \quad (1.118)$$

which from 1.113 gives

$$2d2\pi \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) = 2\pi \quad (1.119)$$

and thus

$$\Delta\nu = \frac{c}{2d} \quad (1.120)$$

where $\Delta\nu/c = 1/\lambda_1 - 1/\lambda_2$. This quantity is called the *Free-Spectral-Range* (FSR) of the etalon.

We can also find the full width at half maximum of each peak. If the phase at the peak is $\delta\phi_p = 2\pi m$, the phase at which the transmitted intensity will drop to half is $\delta\phi_p \pm \delta_{1/2}$, where $\delta_{1/2}$ fulfills the condition

$$\frac{4R}{(1-R)^2} \sin^2(\delta_{1/2}/2) = 1 \quad . \quad (1.121)$$

This in turn means that

$$\delta_{1/2} = 2 \sin^{-1} \left(\frac{(1-R)^2}{4R} \right)^{1/2} \quad (1.122)$$

and thus the total width is

$$\delta = 4 \sin^{-1} \left(\frac{(1-R)^2}{4R} \right)^{1/2}. \quad (1.123)$$

Since R is typically close to 1, the argument of arcsin is small, and we can make a small angle approximation

$$\delta \simeq 4 \left(\frac{(1-R)^2}{4R} \right)^{1/2}. \quad (1.124)$$

In order to convert this into a frequency we use equation 1.113, giving

$$\delta\nu = \left(\frac{c}{2d} \right) \frac{2}{\pi} \left(\frac{(1-R)^2}{4R} \right)^{1/2} \quad (1.125)$$

$$= \frac{\Delta\nu}{F} \quad (1.126)$$

where F is called the *Finesse*. As you see above, the finesse tells us the ratio of the distance between the transmission peaks to their widths. A high finesse therefore means that the peaks are very narrow. The finesse is probably the most frequently quoted number for people discussing Fabry-Perot etalons.

The Fabry-Perot is able to achieve very high resolution, but is not so easy to scan. One way to scan the etalon is to change the distance between the mirrors. However these have to be very well aligned to make sure that the light does not reflect out of side of the etalon. A more popular alternative is to change the pressure of the gas between the etalon mirrors, which changes the refractive index and hence the optical path length. An additional mode of operation for the etalon is to observe the transmission of the etalon as a function of the output angle of the light. From figure 1.20, we can see that in this case the phase difference

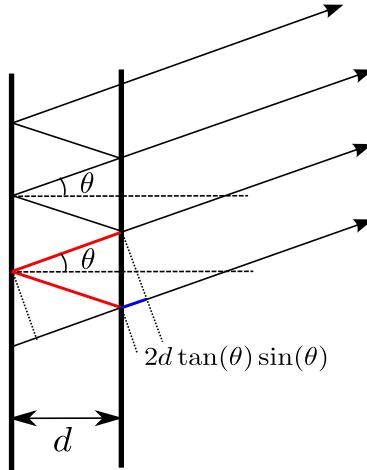


Figure 1.20: Calculation of path difference for parallel rays passing through the Fabry-Perot etalon. The reflected light travels an additional distance of $2d/\cos(\theta)$ between the mirrors of the etalon. Undeflected light travels an additional distance of $2d \tan \theta \sin \theta$ outside of the mirrors.

is given by

$$\delta\phi = \frac{2kd}{\cos(\theta)} - 2kd \tan(\theta) \sin(\theta) \quad (1.127)$$

$$= 2kd \cos(\theta) \quad (1.128)$$

Example rings from a source of Hydrogen and Deuterium are shown in figure 1.21. The two sets of rings are of different intensity, due to the different abundances of Hydrogen and Deuterium in the sample. The separation in frequency is due to the different nuclear masses of the Hydrogen and Deuterium nuclei.

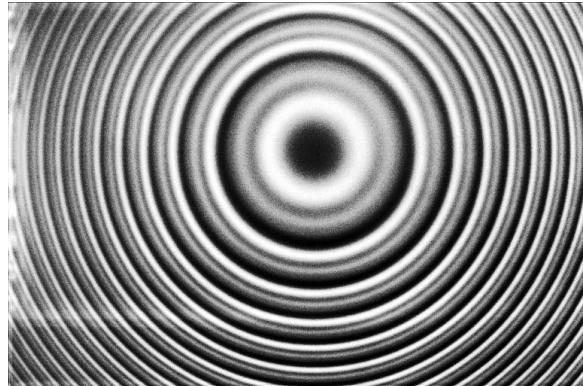


Figure 1.21: Angular fringe pattern of light from a lamp containing both Hydrogen and Deuterium atoms. The abundance of Deuterium is much lower than that of Hydrogen, so the set of Deuterium rings is weaker (these are light grey compared to the whitish Hydrogen rings). The sets of rings are spaced in frequency because although Hydrogen and Deuterium have very similar transitions, the difference in mass of the nuclei leads to small corrections.

Because the etalon is not easy to scan, it is often used in conjunction with another device, such as the grating monochromator. A typical setup is shown in figure 1.22. In this case, the grating selects the region of the spectrum which we wish to observe, and the etalon separates out the fine details.

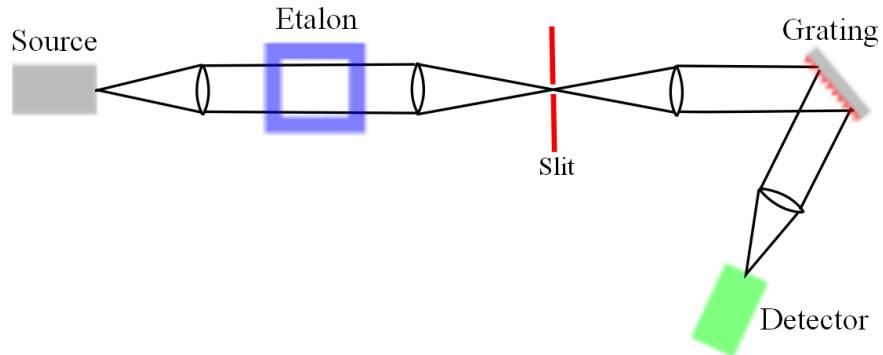


Figure 1.22: Etalon plus grating spectrometer. The grating separates out the lines of interest, and the fine details are uncovered by scanning the etalon.

You should note that reflection co-efficients can be made to be very high indeed. I recently bought an optical cavity with a finesse of > 250000 , which requires $R = 0.999987$. This equates to the light travelling roughly 250000 times around the cavity, or over a path of 50 km, since the cavity is 10 cm in length. Imagine building a Michelson that big! This type of etalon, made from a type of glass which hardly expands or contracts when the temperature changes, is used to stabilize the frequency of lasers for work in quantum information processing and in atomic clocks. The most precise atomic clocks rely heavily on Fabry-Perot etalons.

1.5.5 The use of etalons and gratings in Lasers

Dye Laser: How do lasers relate to the devices which we have been discussing? Figure 1.23 shows a picture of the inside of a commercially built dye laser. One of the most valued properties of laser light is that it is often close to being monochromatic - this feature comes from the cavity which is built around the laser. In the case of the dye laser, the main laser cavity is very long. This means that it has many modes which are very close together in frequency (it has a small free-spectral-range). This causes a problem, because typically we want only the laser to emit one frequency. In order to achieve this, the laser contains a number of frequency selective elements, including very short etalons (with large free-spectral range) and filters which use a dispersive crystal (one whose refractive index depends on wavelength) at Brewster's angle. Some of these are shown in the insets of figure 1.23.

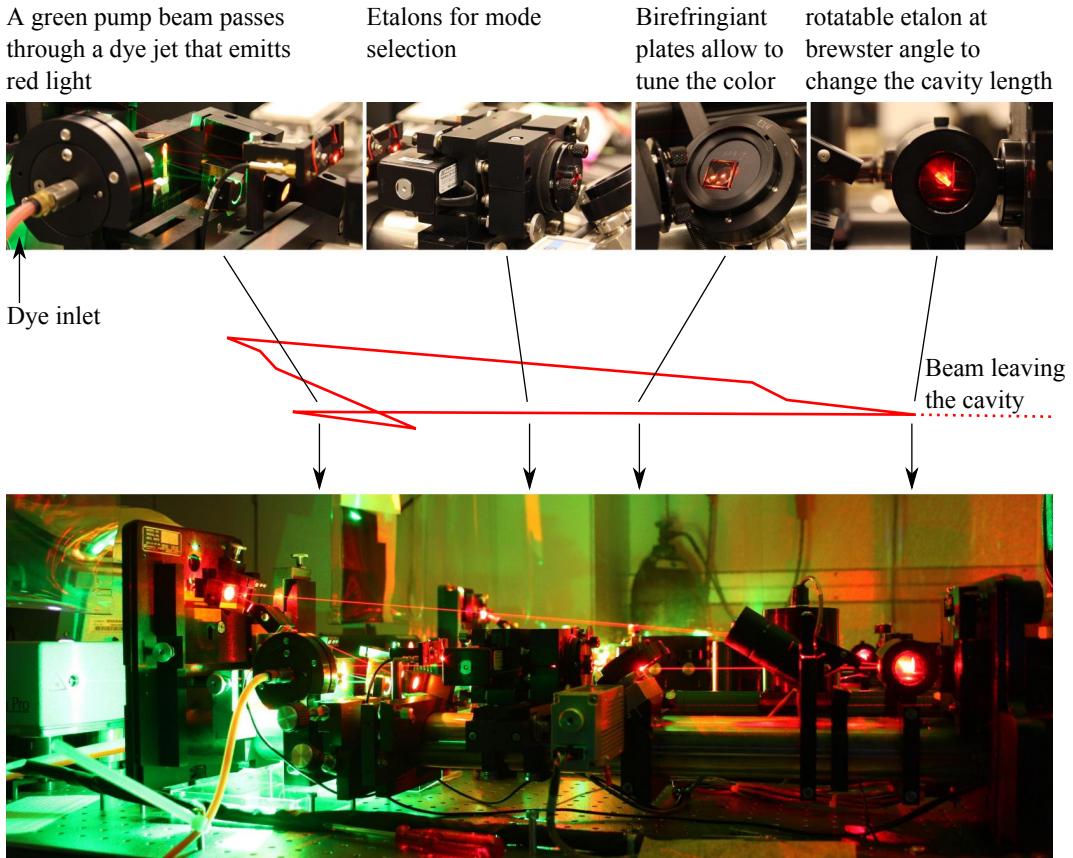


Figure 1.23: Photos from a dye ring laser (F. Merkt group at D-CHAB). The laserbeam circulates in the cavity as sketched and passes several elements as described in the picture.

Diode Laser: Figure 1.24 shows a picture of the inside of a commercially built diode laser. I've pointed out two features, one being the diffraction grating, and the other being the laser cavity. Here the cavity of the laser is formed by the back of the laser diode and the grating itself. The -1 order of the grating sends light back into the diode (a special grating is made which preferentially reflects light into the -1 order). The zero order light from the grating is the light emitted by the laser. Out of the elements in the diode laser, the cavity has the smallest free-spectral range (cavity length is several centimeters) and linewidth, and the grating allows us to pick out one particular wavelength. In addition, the diode itself forms a poor cavity, because it typically has reflective coatings - the diode then forms an additional etalon with a large free-spectral range (because the diode itself is very short), which again adds to the frequency selectivity.

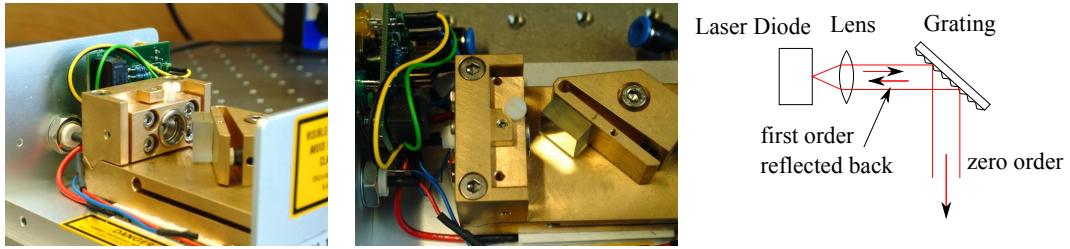


Figure 1.24: A commercial diode laser (Toptica photonics). The diode itself provides the light, and is not visible (it sits inside the metal block on the left). The grating not only provides frequency selectivity, but also forms an element of the optical cavity of the laser, because the -1 order light gets reflected back towards the diode. The light emitted from the laser is the 0 order light from the grating.

1.5.6 LIGO and LISA

A note about one of the largest experiments being performed on earth today. LIGO, the combines the features of both the Michelson and Fabry-Perot interferometers in order to measure gravitational waves. For an understanding of what is being measured, you have to understand General Relativity, so that is far beyond the scope of this course. For us, the important thing is that gravitational waves have an extremely long wavelength, and the experimentalists wish to look at tiny contortions in Space-Time as they pass. To have a chance of observing these they have built a Michelson interferometer with arm lengths d_1 and d_2 of 4 kilometers. Not only that, but they emphasize any change in these arm lengths by using highly reflective optical cavities (such as our Fabry-Perot) in each interferometer arm. The interferometers are sensitive enough that fluctuations of the atom densities on the surfaces of the mirrors become a problem, even after 4 km of arm-length. Tough!

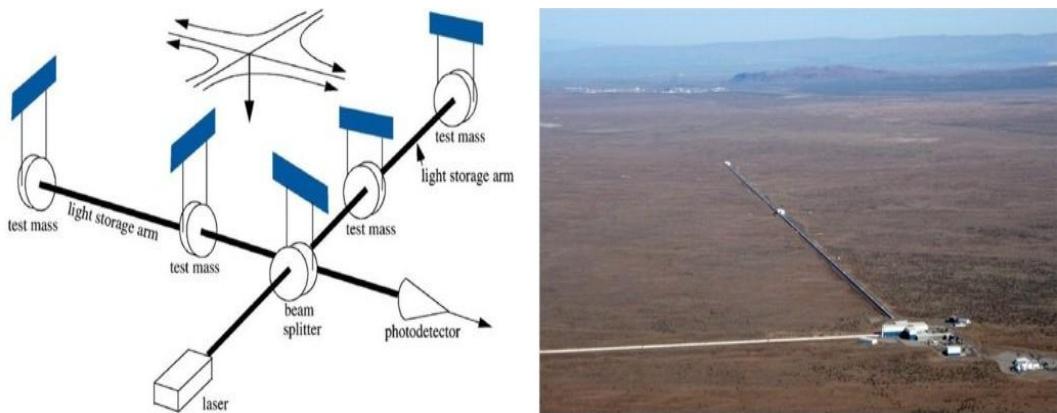


Figure 1.25: The LIGO detector in schematic and a picture of the Hanford site (www.ligo.org). The interferometer consists of two long arms (4 km) that both contain a Fabry-Perot cavity to enhance the photon storage time. The idea is that gravitational waves will change the pathlength in a detectable way.

Of course, physicists want to go further still, and put a large-scale Michelson interferometer into Space. This is the aim of the LISA project. This seems even tougher!

Chapter 2

Photons

In this part of the course we will encounter light as particles called photons, which have momentum and energy. Throughout the 19th century there was an on/off debate about whether light was a wave or a particle. The solutions of Maxwell, and the beautiful work of Huygens, Fresnel and Kirchoff, among others, meant that the view of light as a wave was in the ascendancy throughout the latter part of this period. I should note though, that 100 years earlier Newton was convinced about the “corpuscular” nature of light - that it consisted of particles.

A particle of light is called a photon. It has no mass, but carries momentum, and it travels at the speed of light in all reference frames (you probably encountered this in relativity last term). Here, we will study the experiments which led Einstein to (re-)propose that light is a particle. The basic principles found in these experiments are used a wide range of light detectors. Similar physics allowed people to generate radiation in the X-ray region of the spectrum, which has enabled us to probe the structure of crystals and the internal level structure of atoms in new ways. Thus in this section we are covering a number of important effects for measurement science.

2.1 The Photo-electric effect

The photo-electric effect, as its name suggests, provides a connection between the properties of light (photo) and electricity. As a result, it is widely used in detectors, because electrical signals are the most natural to measure in the laboratory. The effect was first explained by Einstein in 1905, and was actually the work which led to his Nobel prize. Einstein was trying to explain experiments which were performed by Hallwachs in 1888, based on some observations made earlier by Hertz.

The most basic manifestation of the photo-electric effect is the one which was demonstrated in the lecture. A schematic of the experiment is shown in figure 2.1. An metal plate which is initially negatively charged is discharged by incident light from a source. If the plate is initially positively charged, the light has no effect. The introduction of a glass filter which blocked the UV frequencies of the white light source also stops the plate from being discharged. The interpretation is that the charge is carried from the plate by negatively charged electrons which are knocked out of the material by the light.

Quantitative results can be obtained from the apparatus shown in figure 2.2. Here the metal emitter plate is placed in vacuum and a collector plate held at a different voltage is used

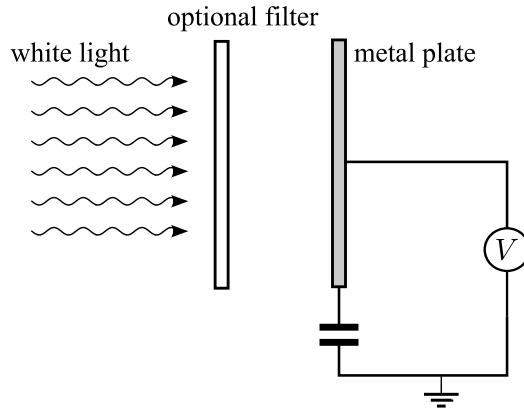


Figure 2.1: Basic setup for observing the photo-electric effect. The light intensity can be varied and a spectral filter can be inserted to block UV light. The metal plate can be charged relative to ground and its voltage relative to ground is measured using a voltmeter. It is found that if the plate is negatively charged, it can be discharged by the light. If it is positively charged, it can not be discharged.

collect the emitted electrons. The important features of the experiment are the following

- A monochromatic light source which can be varied in the amount of power incident on the plate P and the frequency of the light ν .
- A variable voltage source V which allows the potential between the emitter and collector plates to be controlled.
- An ammeter which is used to monitor the current I flowing between the emitter and collector.

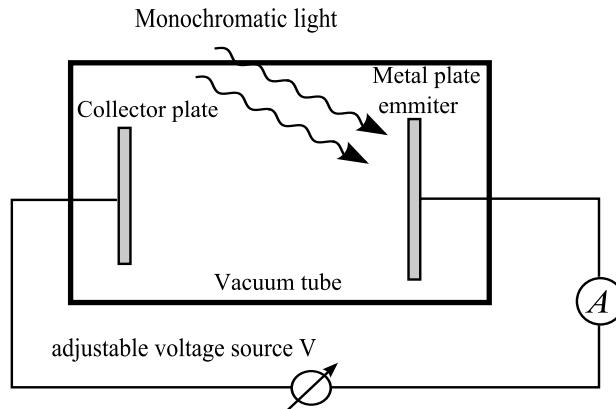


Figure 2.2: Refined setup which allows us to see that the energy of ejected electrons depends on the frequency of incident light.

If light of a certain intensity P and with frequency ν greater than a minimum value ν_{\min} is incident onto the emitter plate, electrons will be ejected. The minimum frequency for which electrons are emitted is characteristic of the material used for the emitter.

By applying a potential difference between the emitter and the collector it is possible to examine the kinetic energy of the electrons which are ejected. If the voltage difference $V = V_c - V_e$ between the collector and the emitter is held at a positive value, the electrons emitted from the emitter plate are accelerated towards the collector. This causes a current

to flow between collector and emitter which is measured on the ammeter. A schematic of the results from such an experiment is shown in figure 2.3. We see that the current saturates as V is increased. For higher incident light power, the saturation current is higher. This implies that the factor which limits the amount of current which flows is the power of the light hitting the emitter.

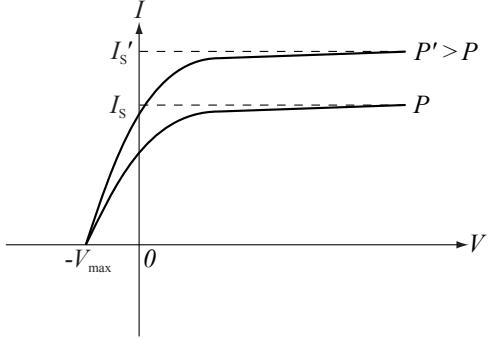


Figure 2.3: Measurement data for the current through the ammeter as a function of voltage applied between collector and emitter. Results are shown for two different light powers. A qualitative description of the results is given in the main text.

If V is held negative, the electrons must climb a potential hill in order to reach the collector plate. At a critical value $V = -V_{\max}$, current no longer flows, because the kinetic energy of the electrons is not high enough for them to climb this potential hill. This allows us to obtain the maximum kinetic energy that the electrons have on leaving the emitter. This maximum kinetic energy is plotted as a function of the frequency of the incident light in figure 2.4.

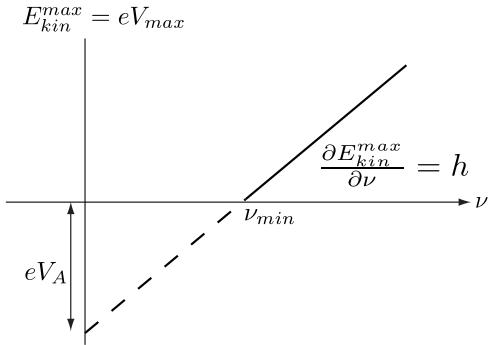


Figure 2.4: The kinetic energy of emitted electrons versus the frequency of the incident light. The minimum energy required to knock an electron out of the metal plate is eV_A . For light with frequency greater than ν_{\min} , the emitted electrons have kinetic energy. The amount of kinetic energy depends linearly on the frequency of the light, and the constant of proportionality is Planck's constant h .

Note that classically these results are unexpected. The electric field of the light would cause the electrons to be accelerated out of the plate, and the field is proportional to the square root of the intensity of the light. One would expect that a larger intensity would increase the field, which would result in higher electron kinetic energies. This is not what is observed in the experiment.

2.1.1 Einstein's explanation

Einstein explained these experiments by envisioning light of frequency ν as arriving in lumps, or “quanta” with energy $h\nu$. We call these *photons*. The number of photons is proportional to the intensity of the light. In order that a photon can knock an electron out of the metal, it must have a certain minimum energy given by $h\nu_{\min}$, which tells us how strongly the electron is bound inside the material. The emitted electron has a maximum possible kinetic energy given by

$$E_{kin}^{max} = h\nu - h\nu_{\min} \quad (2.1)$$

where the constant h is called Planck’s constant, and takes the value

$$h = 6.654 \times 10^{-34} \text{ Js} . \quad (2.2)$$

This constant is named after Max Planck, who had earlier provided an explanation for the spectrum of radiation from blackbody sources by requiring atoms (which he interpreted as harmonic oscillators) to take discrete energy levels separated by $h\nu$. We shall look at the Planck theory later in the course. Einstein’s interpretation of the photo-electric effect extended quantization to light. The value of the Planck constant can be obtained from experimental results of the form shown in figure 2.4.

For a large current to flow between collector and emitter, a large number of electrons must be ejected. This requires a large number of photons. If the light intensity is not high enough these aren’t available, and thus the measured current saturates. Increasing the intensity increases the number of photons available, and thus the saturation current is higher. This explains the principal features of figure 2.3.

Work-function

The amount of energy required to knock an electron out of the metal is called the *work-function* of the material, given by $\Phi = eV_A = h\nu_{\min}$. The work-function is a characteristic of the material. The work-function of a range of materials are given in table 2.1. You should note that these are not the same as the ionization energies of single atoms, which we will see later in the course. Empirically it is found that the work-function is roughly half the ionization energy. For metals (such as Cu, Ag, Al), the electrons are not bound to any one atom, but are shared by many. This is the primary reason why metals conduct electricity and heat. As a result of this, it tends to be easier to extract a single electron from the metal as opposed to individual atoms in a gas. The work-functions depend on many factors, including the crystalline structure and the angle at which light is incident on the surface. Properties related to surfaces include impurities absorbed on the surface and oxide layers (which form readily for many metals). These are also significant factors in determining the work-function. As can be seen from table 2.1, typical work-functions lie in the range between 1 - 6 eV.

2.1.2 Photon Detection

In the preceding lectures, we’ve talked a lot about the detection of light, without saying how this is done. Many of the detectors use the photo-electric effect, so it is appropriate to describe them at this point. One example of a detector which uses these methods is the CCD detector in a camera, another is your eye, which sends electric signals to the brain. One common example in physics laboratories is the photo-multiplier tube. Photo-multipliers are able to amplify the number of electrons emitted for each photon which arrives at the detector, meaning that it is now possible to detect single photons without being limited by

Rb	2.13	Ta	4.19
Cs	1.7 ... 2.14	Mo	4.16 ... 4.2
K	2.25	Cu	4.3 ... 4.5
Na	2.28	Ag	4.05 ... 4.6
Ba	1.8 ... 2.52	W	4.54 ... 4.6
Al	3.0 ... 4.20	Au	4.8 ... 5.4
Zn	4.34	Li	2.2
Pt	5.32 ... 5.66	Ni	5.0

Table 2.1: The work-function of a variety of solid materials given in electron-volts. Al, Zn, Pt, Ta, Mo, Cu, Ag, W, Au and Ni are metals.

electrical noise. I shall describe the photo-multiplier tube here, because we have one as a demonstration in the lecture hall.

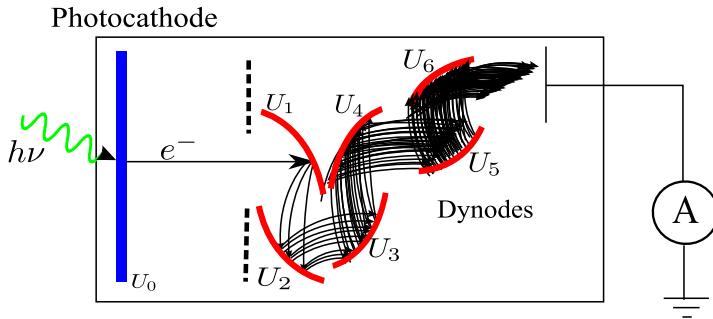


Figure 2.5: The photo-multiplier tube. The incident photon hits the photo-cathode, and ejects one electron. This is then accelerated and focused onto the first dynode, held at voltage $U_1 > U_0$. The extra energy means that multiple electrons are ejected from the dynode. These are accelerated towards the second dynode, which is held at voltage $U_2 > U_1$. This cycle is then repeated for dynodes 3 to 6, with an exponentially increasing number of electrons involved. Finally the amplified signal (now consisting of many electrons) hits a metal plate and travels through a wire to ground. The current of these electrons is measured.

The initial stage of any such detector is a cathode made of some material which has a work-function lower than the energy of the photons that we wish to detect. Thus an incident photon will cause an electron to be emitted from the material. Electrons ejected from the material are then accelerated and focused (by an electromagnetic lens) onto a second plate (called a dynode) which is biased at a higher voltage than the cathode. The electron is accelerated by the potential difference between dynode and cathode, thus it has excess energy, and is able to knock more than one electrode out of the dynode. These electrons are then further accelerated towards a second dynode held at an even higher voltage, and the cycle repeats. The result is that from a single emitted electron, the number of electrons emitted from each dynode increases exponentially until the signal can be reliably measured as a current by electronics.

2.2 X-Rays and Bragg scattering

Electromagnetic radiation spans a huge range of frequencies. Thus far, most of our discussion has been focused on light which has frequencies in the visible range of the spectrum. The wavelength of visible photons is between 400 and 800 nm, so the energies of these photons (given by hc/λ) lie between 1 and 3 eV (an electron-volt, written as eV, is a unit of energy which is equal to $1.60217653 \times 10^{-19}$ J). Now we will turn our attention to the high energy X-ray region of the spectrum, where the photons have energies in the keV range (usually many keV). As a result, the wavelengths are on the angstrom scale, which is the same as the typical spacing of atoms in materials and molecules. This has made X-rays an extremely valuable tool for investigating the structures of materials, primarily in experiments which are related to diffraction.

2.2.1 Production of X-rays

X-rays were first investigated in their own right by Röntgen, who was one of your predecessors here as a student at the ETH. Röntgen was working on experiments which involved the production of high-energy electrons in cathode-ray tubes (see figure 2.6). In a cathode ray tube, a hot filament emits electrons, and these do not pass into the “cathode” but instead this electrode has gaps in it. Electrons passing through the gaps are accelerated down the tube towards the anode by a potential difference applied between anode and cathode. Commonly, the electrons then hit the anode which thus supplies current (the definition of an anode is an electrode to which current must be supplied).

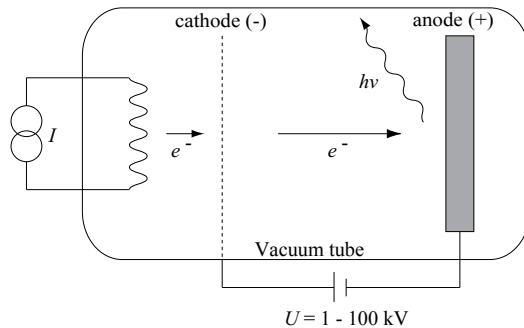


Figure 2.6: Cathode-ray used to generate X-rays. The filament is heated up until thermal electrons are released from the metal. If these have enough energy to get past the cathode, they will start to be accelerated by the potential difference between the anode and cathode.

The cathode-ray tubes in which X-rays were first observed had large voltages applied between its ends (1-100 kV), so the electrons get accelerated to energies of 1-100 keV. Several researchers had noticed funny effects when working with these tubes, such as photographic plates in their laboratories becoming exposed while nominally being kept in the dark. However Röntgen (who wrapped black card around his tube, but still noticed that some radiation was coming out because of fluorescence on nearby cardboard) was the first to make a systematic investigation. The emitted rays were found to be emitted when the electron hit either a metal target or the glass wall of the vacuum tube. X-rays are still often produced by decelerating high-energy (0.1-100 keV) electrons in metal targets. That an accelerated charge releases energy in the form of electromagnetic radiation is a consequence of classical electromagnetism. In the case of X-ray production, this effect is called Bremstrahlung (which any natural German speakers will have more intuition for than me - “braking radiation”). This type of X-ray source is not very pure in frequency, emitting a broad spectrum.

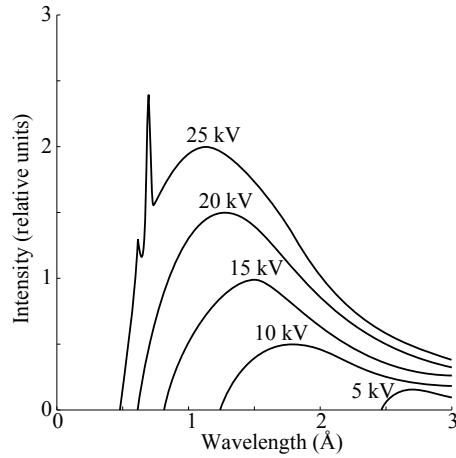


Figure 2.7: Spectrum of electromagnetic radiation generated by decelerating high-energy electrons in a Molybdenum anode. The different curves correspond to date taken with different voltages applied between anode and cathode.

Figure 2.7 shows the X-ray output from a cathode-ray tube in which the accelerated electrons are decelerated in a Molybdenum anode. For increased voltage V difference between the anode and cathode, the intensity of the X-ray radiation rises. For each value of the acceleration voltage there is also a corresponding minimal wavelength of X-ray radiation. Below this wavelength no radiation is generated. This minimal wavelength corresponds to a maximal energy of the emitted X-rays of $h\nu_{\max} = hc/\lambda_{\min}$. In a similar manner to the photo-electric effect, we can find a relation between the voltage applied and the maximum photon energy

$$h\nu_{\max} = eV + \Phi \quad (2.3)$$

The interpretation of this *inverse photo-electric effect* is that the kinetic energy of a single electron is converted to radiation in the form of a photon. Since the electron finishes up bound in the metal the photon can also gain energy from the work-function of the metal. However we should note that the energy of the electrons is typically high compared to typical work-functions (which are in the electron-volt range - see table 2.1) thus the latter can usually be neglected. Radiation is emitted over a continuous spectrum with $\nu < \nu_{\max}$. This indicates that in some cases kinetic energy of each incident electron is converted into multiple emitted photons, rather than just one.

Thus far, we neglected an important feature of the cathode-ray spectra in figure 2.7: for high energies there appear two peaks on top of the bremsstrahlung. These stem from the structure of the atoms that form the anode material. We know this because position of such peaks is a characteristic of the material which is used for the anode. These peaks provide valuable information about the electronic structure of atoms, thus will return to look more carefully them once we have learnt some quantum mechanics.

Because of their high energies X-rays are often used to probe matter where light in the optical domain would be absorbed too easily. The primary medical use was already pioneered by Röntgen, with his wife as the patient. The intensity of the X-rays which passed through her hand is shown in figure 2.8. The X-rays are more likely to be absorbed by the high density material which makes up bones, so we see these in the X-ray image. We also see clearly what seems to be a rather large ring! I do not recommend that you try this experiment at home - some of the negative effects of X-rays are that they ionize atoms in your body (rip off the electrons)- destruction of tissue can occur, which can obviously lead to medical problems.



Figure 2.8: Röntgen's first x-ray picture, taken of his wife's hand in 1895.

2.2.2 Rayleigh scattering

In the next three sections we will look at scattering of photons by atoms. Photon scattering from atoms is a complex subject, which requires some pretty complex quantum mechanics to do properly. Our focus will be to provide simplified descriptions, and see the basic features of experiments. You'll see that some of the language will start to get mixed up between waves and particles. This is unavoidable, and will be one of the common themes from now until the end of the course. First up, we'll look at scattering for which the wave picture is satisfactory, and for which the classical explanation words well. It is a type of scattering which we observe every day, because it is the reason that the sky appears blue.

When electro-magnetic radiation is scattered *elastically* by particles, we talk about *Rayleigh scattering*. It is a process that can be described classically. Imagine an oscillating electric field of amplitude E and frequency ω which is incident on a bound electron. Any binding potential can be expanded in a power series and for small excitations we can keep the first terms only. Therefore we can treat the bound electron as a classical harmonic oscillator, with natural oscillation frequency ω_0 . The electric field of the radiation excites the motion of the electron, which then oscillates according to

$$m \frac{d^2x}{dt^2} + \frac{1}{m} \omega_0^2 x = -eE \cos \omega t \quad (2.4)$$

Inserting a solution $x = x_0 \cos \omega t$ into this expression, we find that the resulting amplitude of the driven oscillator is

$$x_0 = \frac{eE}{m(\omega^2 - \omega_0^2)}. \quad (2.5)$$

From your study of the Hertz dipole last year (equation 11.111 in Rainer Wallny's lecture notes), you may vaguely remember that the acceleration of the charge distribution causes the electron to emit radiation. The electric field amplitude is

$$E_{\text{emit}} \propto \frac{d^2x}{dt^2} \propto \frac{eE\omega^2}{m(\omega^2 - \omega_0^2)} \cos \omega t \quad (2.6)$$

The emitted light thus has the same frequency as the incident light, and thus the outgoing photons must have the same energy as the incident photons. This is why we call this elastic

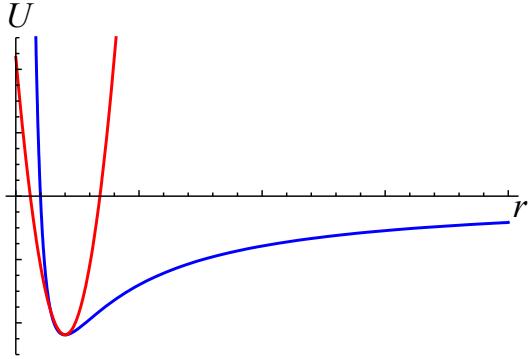


Figure 2.9: Binding potential an electron sees from its atom and a quadratic fit at the potential minimum.

scattering of light. This approach is valid in the limit that the photons don't have high energy. If they did, the electron would be ripped out of the atom by the incident field, and we could not treat it as a harmonic oscillator. Note the dependence of the scattering rate on ω . In air, we have Oxygen and Nitrogen molecules, which have resonant frequencies deep into the ultra-violet range of the spectrum (between 10 and 300 nm). This means that Rayleigh scattering is stronger for blue light than for red light, which is why you see that the sky is is blue.

2.2.3 Bragg scattering

Bragg-scattering of X-rays occurs in crystalline materials, where atoms are held in regular arrays. If radiation is incident on the crystal, then the electrons of all the atoms will be excited with a well defined phase relative to each other. They will re-radiate the “scattered” radiation by Rayleigh scattering. Let us imagine that the light is radiated as a spherical wave. Because the crystal has a regular array, there is interference between the light which is scattered from each of the atoms. Figure 2.10 gives a sketch. The phase difference between light scattering from neighboring crystal planes is

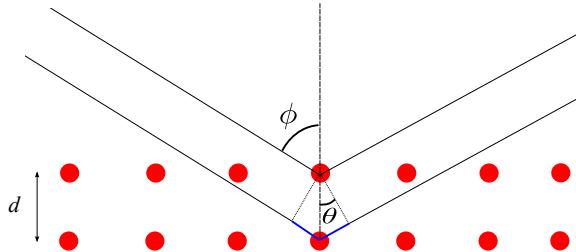


Figure 2.10: Bragg scattering. The incident light is Rayleigh scattered by each individual atom, which thus emits a spherical wave. Interference occurs at angles where the difference in paths between light scattered off atoms in each layer (layer spacing d) of the lattice is a multiple of 2π .

$$\delta\phi = \frac{2\pi}{\lambda} 2d \sin(\theta) = \frac{2\pi}{\lambda} 2d \cos(\phi) \quad (2.7)$$

We get constructive interference if this phase shift is equal to 2π , therefore for

$$2d \cos(\phi) = \lambda \quad (2.8)$$

This condition allows experimenters to deduce the structure of crystals. There are a number of different experimental setups which are used for this. One is shown in figure 2.11. X-rays from a polychromatic source (many frequencies) are scattered from a crystalline target, and then strike a photographic plate. The Bragg scattering leads to interference in particular directions, which are observed on a phosphorescent screen. In another other common method a monochromatic X-ray source is used, and the crystal under study is mounted on a rotatable mount, with the X-rays scattered measured as a function of the angle of the crystal both to the incoming X-rays and the angle along which the detector is placed.

The interference pattern observed in Bragg scattering experiments is related to the Fourier transform of the crystal structure, and this structure can be obtained from the data using an inverse Fourier transform. You will likely learn more about this in courses on condensed matter physics. Though we have illustrated the basic technique of Bragg scattering using a single set of parallel planes, actually in real crystals there are many possible planes of parallel atoms. An example of different planes is shown for a square lattice in figure 2.12. You should note at this point that for Bragg scattering to occur $\lambda \leq 2d$. In crystals, d is typically on the order of 1 nm, therefore the required wavelengths fall in the X-ray region of the spectrum.

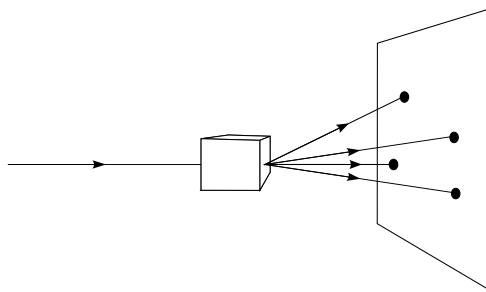


Figure 2.11: Schematical setup for a Bragg scattering experiment. A polychromatic X-ray beam is sent through a crystal. On a screen diffraction at certain wavelengths can be detected.

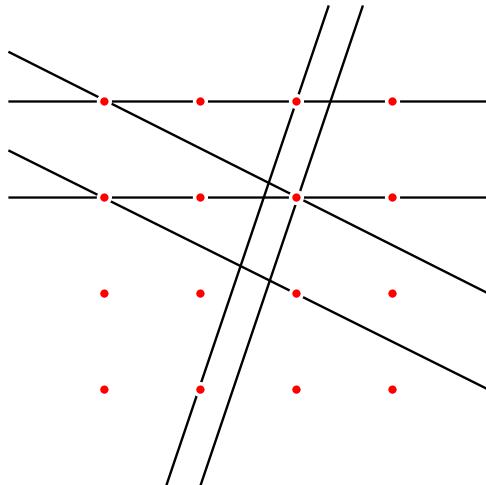


Figure 2.12: A crystal which has atoms on the corners of a periodic square lattice has multiple planes from which Bragg scattering can occur.

Bragg scattering not only allows for the spatial structure of crystals to be reconstructed, but also gives a method for separating out spectral components of the X-ray radiation. The usual devices we use in optical spectroscopy are not able to do this easily. This is because

X-rays have a refractive index close to 1 for most materials (you know this already because X-rays pass through your hand) and thus there is a lack of lenses and mirrors at these wavelengths. Instead, the X-rays are scattered from a crystal with well known structure. Since the spatial structure is now known, the angles at which Bragg peaks appear gives information about the wavelength of the X-rays. The one device which is carried over from optical spectroscopy is the grating spectrometer. These must be used at very small incidence angles (thus providing an effective small slit separation as viewed from the incoming X-ray direction) because the wavelength of the X-ray radiation is so small.

2.3 Photon momenta – The Compton effect

Further evidence for properties of photons comes from observation of the Compton effect. Here light is scattered by electrons but in contrast to Rayleigh scattering, in this case the frequency of the scattered light is changed. The shift is independent of the material used, but does depend on the direction in which the scattered photons are detected. To explain his observations Compton had to treat the light as particles and assign a momentum of $p_\nu = E_\nu/c$ to them. If a photon is scattered by an electron and the binding energy of the electron is very small compared to the energy of the incident photon, then effectively we can ignore the binding completely, and treat the electron as a free particle, which gains kinetic energy when the photon is scattered. This is what happens in Compton scattering. Since the binding energy is negligible, the effect is almost independent of the material.

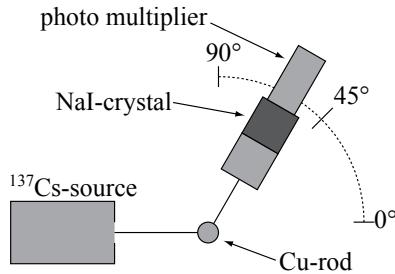


Figure 2.13: Experimental setup used to demonstrate the Compton effect in the lecture. High-energy γ -ray photons are emitted from a Cesium source and scattered by the electrons in a copper rod. The NaI crystal is used as a scintillation counter, releasing an amount of light which is proportional to the energy of the incident photon. This light is detected on a photo multiplier tube. The detection apparatus is mounted on a moveable arm, which allows the spectrum to be recorded at a range of angles.

Compton scattering occurs when extremely high photon energies are used. Photons of energy $E_i = h\nu$ are incident on a target, and outgoing photons are detected at an angle θ to the incident beam. The spectrum of these outgoing photons is measured as a function of the angle. These X-rays are not close to the Bragg condition, so we don't have to worry about Bragg scattering. Compton observed at a scattering angle of 90° that the spectrum is frequency shifted with respect to the incident photons. His original data is shown in figure 2.14. Compton determined the spectrum of the light using Bragg reflection from a calcite crystal. By measuring the spectrum of the X-ray radiation at 0 and 90° and comparing the features of two spectra, he was able to deduce a wavelength shift of 0.022 Å.

In the demonstration experiment we looked at the spectrum of the scattered γ -rays (which have even higher energy than X-rays) at different detection angles. We obtained the spectrum using a scintillation counter, rather than by Bragg scattering. In this device, the γ ray knocks an electron out of the material, which then passes through a phosphorescent material, causing it in turn to emit light. This light is then detected on a PMT. The kinetic energy of the electron determines the number of photons emitted by the phosphorescent material. Thus the amount of light detected at the PMT is related to the energy of the incoming γ radiation. The frequency shift as a function of observation angle obtained such an experiment is shown in figure 2.15.

Compton scattering involves X and γ -rays which have energies similar to the mass energy of the electrons $m_e c^2$ from which they scatter, and much larger than the binding energy of the electrons in the material. Therefore in treating Compton scattering we must use a relativistic approach, and can neglect the binding energy. Figure 2.16 shows the relevant parameters for the scattering process. We will assume the electron is initially at rest, so its initial energy is purely mass energy $m_e c^2$, and its initial momentum is zero. The photon, at initial frequency ν , has an initial momentum given by $p_\nu = h\nu/c$. After the collision, the

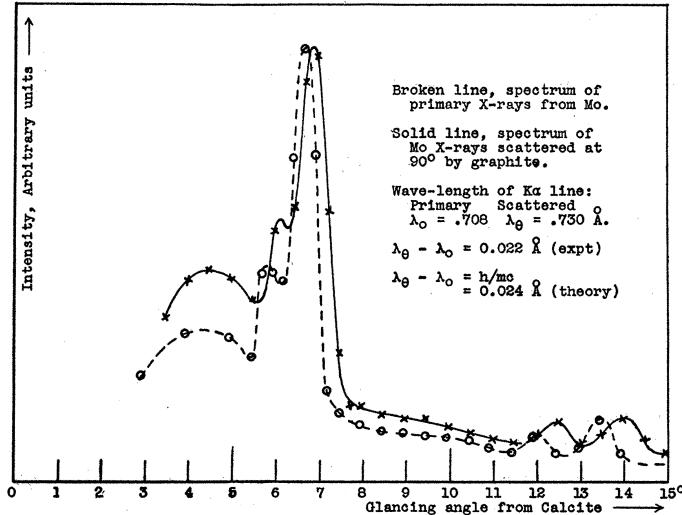


Figure 2.14: Compton's original measurement data. He performed his measurement using a X-ray tube with a molybdenum anode and a piece of graphite for the scattering. In Compton's experiment the wavelength of the light was measured using Bragg reflection from a Calcite crystal. By measuring the spectrum of the X-ray radiation at 0 and 90° and comparing the features of two spectra, he was able to deduce a wavelength shift of 0.022 Å. The specific form of the curve is due to the spectrum of the X-rays generated by the molybdenum anode. Source: A.H. Compton, Phys. Rev. 21, 483 (1923)

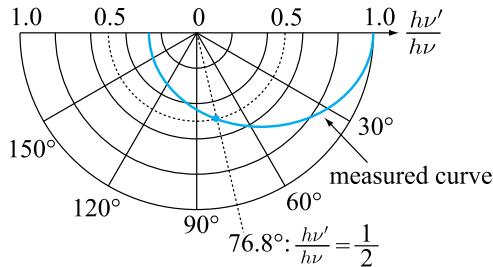


Figure 2.15: Relative frequency shift of scattered photons, plotted for different detection angles.

photon energy and momentum are shifted to $h\nu'$ and $h\nu'/c$ respectively, and the electron has E_e and momentum $\mathbf{p} = (p_x, p_y)$. By conservation of momentum

$$\begin{aligned} 0 &= p_x + p_{\nu',x} = p \sin \phi - \frac{h\nu'}{c} \sin \theta \\ p_{\nu,y} &= p_y + p_{\nu',y} \\ \frac{h\nu}{c} &= p \cos \phi + \frac{h\nu'}{c} \cos \theta \end{aligned} \quad (2.9)$$

where $p_{\nu,x}$ is the x component of the photon momentum for the incident photon, $p_{\nu',x}$ is the x momentum of the outgoing photon and p_x is the x momentum of the electron after the scattering has taken place. The magnitude of the final momentum of the electron is p .

Energy conservation requires that

$$h\nu + m_e c^2 = h\nu' + (m_e^2 c^4 + p^2 c^2)^{1/2} \quad (2.10)$$

and thus

$$\begin{aligned} m_e^2 c^4 + p^2 c^2 &= (h(\nu - \nu') + m_e c^2)^2 \\ p^2 c^2 &= (h\delta\nu)^2 + 2h\delta\nu m_e c^2 \end{aligned} \quad (2.11)$$

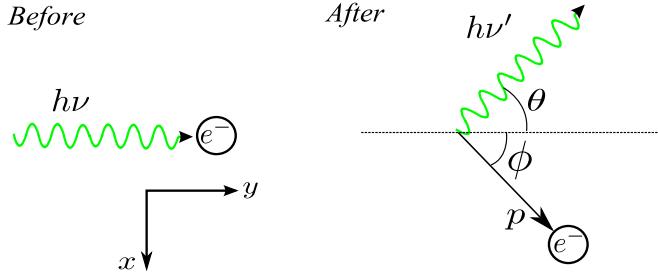


Figure 2.16: Compton scattering. An incident X-ray photon with frequency ν scatters off an initially stationary electron. The photon energy is so high that the electron binding energy is negligible. After the collision, the photon has a different frequency, and the electron gains momentum \mathbf{p} .

Where we abbreviated $\delta\nu = \nu - \nu'$. From the equations for momentum conservation, we can find

$$p^2 c^2 = (h\nu')^2 \sin^2 \theta + (h\nu - h\nu' \cos \theta)^2 \quad (2.12)$$

$$\begin{aligned} &= (h\nu')^2 + (h\nu)^2 - 2h^2 \nu \nu' \cos \theta \\ &= (h\delta\nu)^2 + 2h^2 \nu \nu' (1 - \cos \theta) \end{aligned} \quad (2.13)$$

Equating the two expressions for $p^2 c^2$ we find that

$$\delta\nu = \frac{h\nu(\nu - \delta\nu)}{m_e c^2} (1 - \cos \theta) \quad (2.14)$$

which along with $\delta\lambda = c/\nu - c/(\nu - \delta\nu) = c\delta\nu/(\nu(\nu - \delta\nu))$ gives

$$\delta\lambda = \frac{h}{m_e c} (1 - \cos \theta) = \lambda_c (1 - \cos \theta) \quad (2.15)$$

where λ_c is known as the Compton wavelength. This expression tells us that the wavelength shift is fixed for a given angle – it is given purely by a constant which consists only of the electron mass, Planck's constant, and the speed of light. The agreement between theory and experiment provided convincing evidence that the photon carries momentum.

Chapter 3

Atoms and their composition

Much of matter is made up of atoms. An atom of a given type is a unit involving three types of particles bound together, and the unit is found to have the same properties whether we look at it here, or in outer space. However, if we were to wind back 110 years, we would find that there was a lively debate going on as to whether atoms existed at all. This was tied up with arguments about the nature of light, and whether “stuff” goes around as lots of discrete lumps or as something else.

Of course, as with many objects in physics, there are things we want to know, such as its size, mass and structure. This is as true for the atom as for anything else, and much of the experimental work at the beginning of the last century went into trying to tie down these properties. The methods which were first used to measure these properties are of much modern relevance, since they form the basis for many experimental methods in physics today. For example, early experiments looked at atomic structure by scattering electrons (and alpha particles - a Helium atom with no electrons) from matter. Today, the collisions produced at particle accelerators like CERN are again looking at the results of scattering particles from one another, albeit at far higher energies than could have been imagined in the early days of the last century. A second example involves the methods used to find the relative mass of atoms. Mass spectrometry has now been refined and is heavily used in chemistry and biology for deducing the products of chemical reactions.

3.1 Introduction to atoms

Atoms are made up of protons, neutrons and electrons. For the purposes of this course, we will take these to be fundamental units. We now know that protons and neutrons have structure, but it requires extremely high energies to observe this because the binding between the constituents is very strong. It is for this reason that we will ignore this here. Both the proton and neutron have masses close to 1.66×10^{-27} kg. The proton has a charge of 1.602×10^{-19} Coulomb, whereas the neutron is neutral. The electron has the opposite charge to the proton, but is much lighter, having a mass of $9.10938188 \times 10^{-31}$ kg.

3.1.1 Periodic table of elements

In 1869 Dmitri Iwanowitsch Mendeleev and Lothar Meyer independently drew up a table of the elements which organised them in order of increasing mass and in columns according to their chemical and physical properties. It is shown in figure 3.1. Figures 3.2 and 3.3 show

how atomic radii and ionization energy display a repetitive pattern dependant on the atomic number Z . Today it is known that atoms consist of a charged nucleus (made up of protons and neutrons) that is surrounded by electrons. Z gives the nuclear charge (expressed in units of the proton charge), which for neutral atoms is equal to the number of electrons.

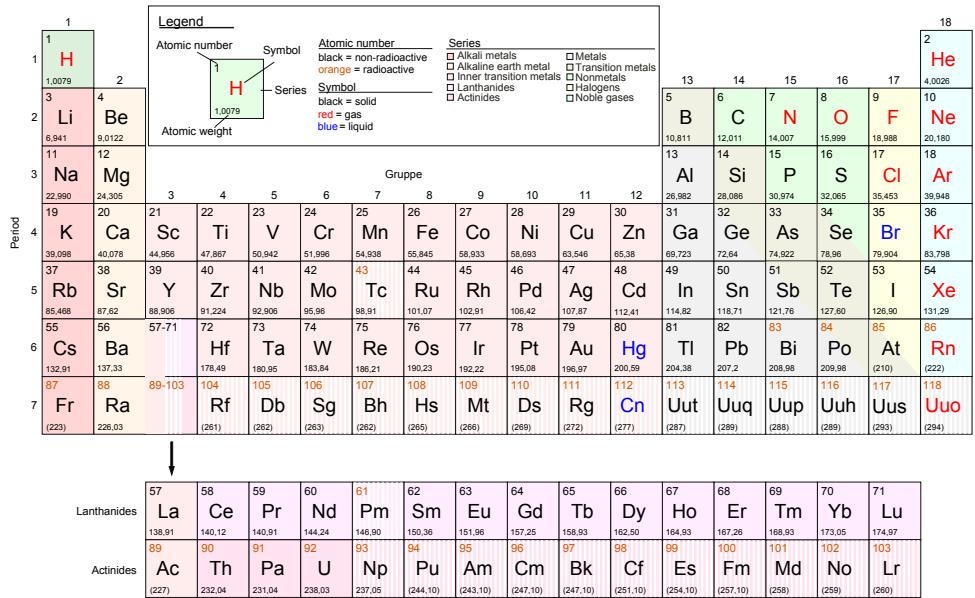


Figure 3.1: The periodic table of the elements. Each element is labeled by a letter representing the name, the number of protons which are in the nucleus, and the mass of the element, which is averaged over the natural abundances of the different isotopes (see section 3.2.2). The columns of the table are aligned according to common chemical properties.

The number of electrons defines the most important properties of any element, because electrons occupy a larger space and therefore interact with the outside world more strongly. The electronic structure is thus responsible for most of the atom's chemical properties. Mendeleev and Meyer preceded any explanation of the reason for the periodic behavior which they classified in the periodic table by at least 50 years. With the onset of quantum mechanics, it was possible to explain the features of atoms, and of the ordering in the table. We will learn more about this later in the course.

Each element is given a name, many of which were made up long before the notion of atoms came along. It is common to denote an atom by the following form

$${}^A_Z G \quad (3.1)$$

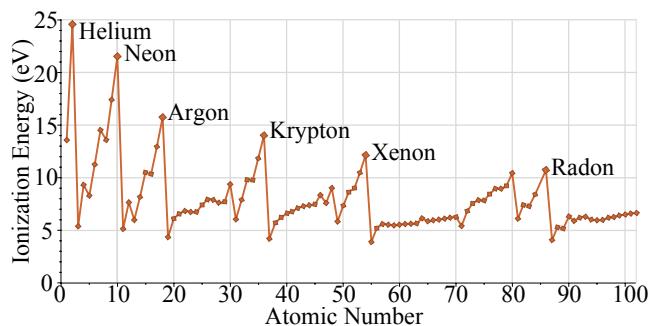


Figure 3.2: The energy needed to ionize different atoms. There are different ways to strip an atom off one electron, e.g. by electron bombardment or with laser light of a high enough energy.

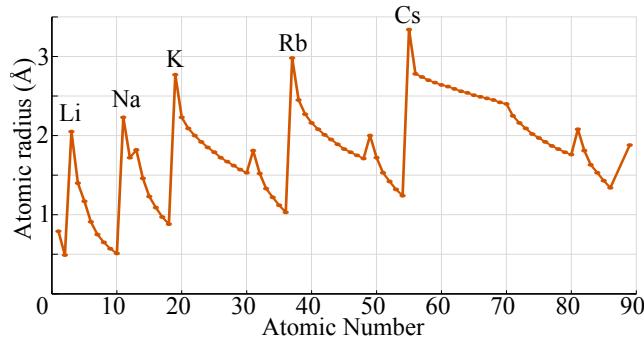


Figure 3.3: Atomic radii in dependence of the atomic number. As a measure for the radius their distance in solids or molecules is used.

where G is the element name, A is the mass of the atom (in units of “atomic mass”, see below) and Z is the charge of the nucleus. You might think this notation is a bit redundant from what we have just discussed, but the reason for it will become clear as the lectures progress. Some of you may already have spotted that Z gives the number of protons, and the number of neutrons must be $A - Z$, because the electron mass is so much less than the mass of the proton and neutron. You may also notice that in the periodic table of figure 3.1, some of the masses are not integer multiples. We shall come back to this later in the lecture.

3.2 Measurements of atomic mass

A first insight into the mass of atoms comes from experiments with gases. From thermodynamics, you already know that under the same conditions, ie. for a gas at the same pressure in the same volume, there will always be the same amount of particles. This is summarized in the equation of state for an ideal gas (an approximation consisting of point particles which do not interact with each other), which takes the form

$$PV = Nk_B T \quad (3.2)$$

where P is the pressure, V the volume, T the temperature and N the number of atoms. Note that though this equation works pretty well for dilute gases, it has nothing to do with the mass of the particles. The equation does tell us that we can make an estimate of the particle number from the pressure, volume and temperature of the gas. Thus by weighing samples of different gases with the same temperature, pressure and volume, we can determine the weights of the constituent atoms.

Experimental results from this type of measurement indicate that the weight of each type of atom is close to an integer multiple of the weight of the hydrogen atom. Therefore by using the hydrogen mass “u” as the unit of mass, we can give all other atomic masses relative to it, e.g. $A_{\text{rel}}(\text{O}) \approx 16$ u or $A_{\text{rel}}(\text{Ca}) \approx 40$ u. It turns out that hydrogen is a particularly difficult atom to measure the weight of. It is highly reactive, and very light. So more precise measurements can be made of higher mass atoms. The “standard” is carbon. Instead of defining the atomic mass unit “u” as the hydrogen mass, we say that the atomic mass unit

$$1 \text{ u} = 1/12 \text{ of the mass of } {}_6^{12}\text{C} . \quad (3.3)$$

Thus we have determined the mass of samples of atoms relative to another. To find the absolute mass of a single atom, we have to be able to count the number of atoms in a sample of a given weight. This is extremely hard to do exactly (the numbers are very large), but there are methods which allow us to obtain estimates. One possible method uses Bragg scattering of X-rays, which allows us to find the distance between planes of atoms in a crystal. If this can be done for all crystal planes (crystals are rarely completely isotropic) it is possible to obtain the number of atoms in a volume. The mass of a single atom is then easily obtained

$$M_{\text{atom}} = \frac{\text{Total mass of a sample}}{\text{Number of atoms in its volume}} \quad (3.4)$$

Example masses for some common elements are given below.

$$\begin{aligned} 1 \text{ u} &= (1.6660565 \pm 0.0000005) \times 10^{-27} \text{ kg} \\ M({}^1\text{H}) &= 1.67342 \times 10^{-27} \text{ kg} \\ M({}^{12}\text{C}) &= 19.92516 \times 10^{-27} \text{ kg} \\ M({}^{16}\text{O}) &= 26.5584 \times 10^{-27} \text{ kg} \end{aligned}$$

As well as the method described above, there are a number of other ways of measuring the number of atoms in a sample. Others include

- Electrolysis (the current can be measured and the mass change observed)
- Gas constant and Boltzmann’s constant
- Radioactive decay (the loss of mass as a function of time can be related to the number of decay events)

3.2.1 Mass spectrometry

Though the methods used above are able to measure the average mass of atoms in a sample, they make the assumption that all the atoms are the same. In many gases, we find a mixture of elements, and it is useful to be able to find the masses of all the constituents. The devices we use to do this are called mass spectrometers, since they provide a “spectrum” of masses.

Mass spectrometers make use of the fact that electric and magnetic fields can be used to deflect charged particles. This allows the separation of particles with a different mass to charge ratio m/q . The most simple (and oldest) arrangement of this type is due to Thomson (1913). The method uses a beam of ions which are created from neutral atoms in a gas discharge (by heating the atoms up, some of them collide with each other, electrons or the walls of the container in such a way as to knock an electron out of the atom). These ions are then passed through a short region with parallel electric and magnetic fields, both perpendicular to the initial velocity of the ions. The basic arrangement is shown in figure 3.4.

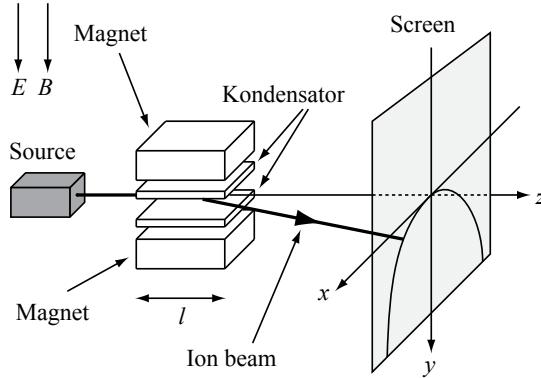


Figure 3.4: A mass spectrometer which uses the parabola method consists of a source that produces an ion beam, an arrangement of electric and magnetic fields to deflect the beam and a screen.

An ion of charge q and mass m passing along the z -axis through an electric field E pointing in y -direction will feel a force which accelerates it according to

$$\ddot{y} = \frac{q}{m} E. \quad (3.5)$$

Integrating this expression twice, and taking initial conditions $y = 0$, $\dot{y} = 0$, we find that for a region of electric field of z -thickness l

$$y = \frac{1}{2} \frac{q}{m} Et^2 = \frac{1}{2} \frac{q}{m} E \frac{l^2}{v_z^2}, \quad (3.6)$$

where the second expression follows from replacing the time t that the particle spends in the field with l/v_z and the particle's z -velocity, v_z . In the presence of the additional B field, this is an approximate expression, since v_z would be expected to change in time. The approximation is good if the deflection due to the B -field is not too large.

The B -field induces a Lorentz force

$$\mathbf{F}_L = q\mathbf{v}_z \times \mathbf{B} \quad (3.7)$$

which acts perpendicular to the particle velocity. In the absence of the electric field, it would give rise to circular motion in the $x - z$ plane, the radius of which can be found by equating

the Lorentz force to the centripetal force $F_c = mv_z^2/r$. We find

$$r = \frac{mv_z}{qB} \quad (3.8)$$

with an initial centripetal acceleration of

$$a_c = \frac{v_z^2}{r} = \frac{q}{m} B v_z. \quad (3.9)$$

If the region of the E and B fields is short enough that the z -component of the velocity is not significantly changed, its x -acceleration will be roughly constant, and equal to this centripetal acceleration. We can then calculate the deflection in x -direction

$$x = \frac{1}{2} a_c t^2 = \frac{1}{2} \frac{q}{m} \frac{Bl^2}{v_z}. \quad (3.10)$$

Note that x goes as the reciprocal of v_z , whereas y goes as the $1/v_z^2$. If many particles with different velocities are passed through the apparatus, and the output is collected on a screen placed immediately after the electric and magnetic field region (note that this is not the position of the screen in figure 3.4, which is some distance behind this region), then each velocity will correspond to a particular pair of co-ordinates x and y . We can see the relationship between the two immediately after leaving the field region by combining equations 3.6 and 3.10 (and eliminating the velocity v_z) yielding

$$y = \frac{2E}{l^2 B^2} \frac{m}{q} x^2. \quad (3.11)$$

This is the equation of a parabola. Examples of experimental results can be seen in figure 3.5. Since E , B and l are known, their shape allows us to obtain the mass-to-charge ratio. An initial ion beam containing different isotopes of a single atomic species will produce a different parabola for each isotope. The relative intensities of the parabolas give the relative abundance of the isotopes in the probe.

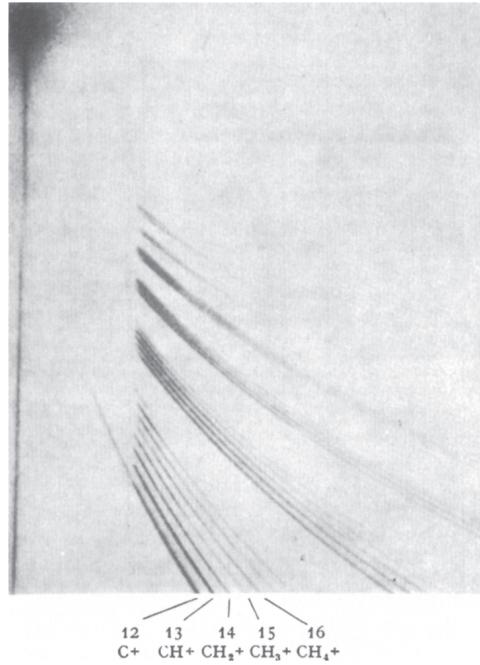


Figure 3.5: Parabolic traces of a mixture of hydrocarbon ions seen on a screen. Photo after Conrad from W. Finkenburg: Einführung in die Atomphysik.

One problem with the parabola method is that the ions with a given m/q but with different initial velocities are spread over the whole parabola. This means that the signal is spread

out over the screen, which reduces sensitivity. A better approach, as with optics, is to focus the signal to a single point, where in this case this means that all ions with the same m/q end up at the same place on the detector. Modern mass spectrometers use more complex field configurations to do this, allowing the detection of faint traces of particles and allowing particles of different masses to be separated with a high resolution.

Today mass spectrometers are used for precision measurements in atomic and nuclear physics, for chemical analysis, to detect residual gases in vacuum systems and to provide pure molecular or atomic beams.

3.2.2 Isotopes

In most of the atoms in the periodic table, the atomic mass (in units of u) is roughly two times the atomic number. This is because the most stable nuclei are formed with similar numbers of protons and neutrons (neutral mass particles of similar mass to a proton). While the periodic table gives only one value of A for a given Z , actually most atoms have multiple *isotopes*, that is variations with different mass, but the same nuclear charge. Since these will have the same number of electrons, they have very similar electrical properties to each other, however in mass-spectra it is possible to see the different masses. Since the atomic mass changes but the charge does not, isotopes differ in having different numbers of *neutrons* in their nucleus.

Actually, the atomic weight given in the periodic table is a weighted average, with the mass of each isotope multiplied by its natural abundance. As noted before, atoms with equal numbers of protons and neutrons have more stable nuclei. They are less likely to undergo nuclear decay and therefore have much higher natural abundance. For example, in calcium the most common isotope is $^{40}_{20}\text{Ca}$, which has a natural abundance of 96.94 %. By contrast, $^{43}_{20}\text{Ca}$ has a natural abundance of 0.135 %. In nuclear physics, it is often useful to use isotopes which are not stable (these undergo nuclear reactions easily). It is then desirable to separate out different isotopes, to isolate those which are useful for these purposes. This is performed using mass spectrometers. Unfortunately, one purpose of isolating particular isotopes is to create nuclear armaments – trade in certain types of mass spectrometer has thus become very carefully monitored.

3.3 Size of the atom - scattering experiments

In a game of snooker/billiards/pool, the probability that two balls collide depends on their size. Therefore it is natural to try to determine atomic size by scattering other particles from them. A sketch of one setup for such an experiment is given in figure 3.6. Particles from a source are incident on a sample (typically either a gas or a thin foil). On the other side of the sample, a detector allows us to record the number of particles which pass through the sample without being significantly deflected.

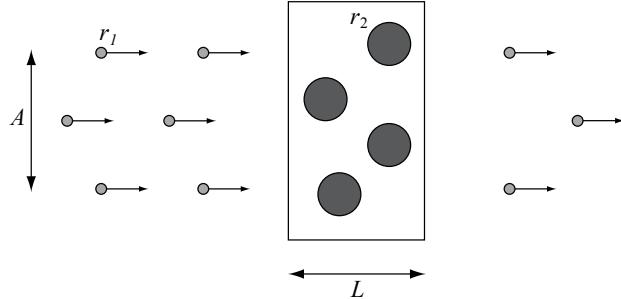


Figure 3.6: Schematic setup for a scattering experiment. Particles of radius r_1 are incident on an area A of a target of length L containing particles of radius r_2 . The particles passing the target undeflected are detected.

A first approximation to particle scattering experiments which is useful in terms of the concepts involved is to think of both the particles of the incident beam and in the sample as hard spheres (as for snooker balls). This is a good approximation for scattering experiments with neutral atoms, because they don't produce any long-range repulsion such as that due to the Coulomb interaction. If the radius of the incoming particles is denoted as r_1 and that of the particles in the sample r_2 , then an incoming particle will be deflected if its trajectory brings it closer to a particle in the sample than the sum of the two radii. Therefore each particle in the sample has an associated region of area $\sigma = \pi(r_1 + r_2)^2$ that particles cannot pass through without being deflected. We call this region σ the *cross section* for scattering of one particle type by the other. Use of cross-sections extends far beyond hard-sphere models, as we shall see later on in the section on Rutherford scattering.

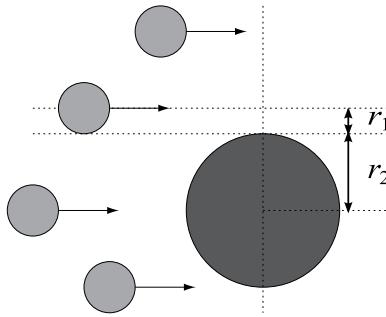


Figure 3.7: If the trajectory of two particles gets closer than $r_1 + r_2$ they will be scattered.

The cross-section for an incoming particle to be scattered by a sample containing multiple particles is the sum of the cross-sections for all of the particles in the sample, so long as we work in the approximation that none of the particles in the sample shadow each other (ie. that none of the atoms sit behind another and thus can never be reached by an incident particle). Let us also assume that the incident beam of particles is evenly distributed over an area A , which is also the area of the sample. We then find that the probability that an

incident particle is scattered is

$$\begin{aligned} P_{\text{scatt}} &= \frac{\text{Area of all interaction cross - sections in the volume } V \text{ of the beam}}{\text{Total area A}} \\ &= \frac{\sum_{i \in V} \sigma_i}{A}. \end{aligned} \quad (3.12)$$

The approximation is valid in dilute gases which aren't too thick, or in very thin solid films.

Now consider a situation where we have N particles incident on a sample. The change in the number of particles in the beam due to scattering out of the beam is given by

$$\Delta N = -P_{\text{scatt}}N = -\frac{\sum_{i \in V} \sigma_i}{A}N. \quad (3.13)$$

If the cross section for all scatterers is σ and the density of particles in the *sample* is n , the numerator can be written as $\sum_{i \in V} \sigma_i = nV\sigma$. Now let us slice the volume of the sample into small pieces of length dx that have a volume $V = Adx$. The number lost from the beam due to scattering in each slice dN is

$$dN = -\frac{n\sigma Adx}{A}N. \quad (3.14)$$

Integrating over several slices up to a total thickness x , we find that the number of particles which are still in the beam is

$$N(x) = N(0)e^{-n\sigma x}. \quad (3.15)$$

where $N(0)$ is the initial number of particles in the beam. Therefore, after a sample of length L , $N(L) = N(0)\exp(-n\sigma L)$ particles are still in the beam, and accordingly

$$N_{\text{scatt}} = N(0)(1 - e^{-n\sigma L}) \quad (3.16)$$

will have been deflected.

We now have all the relations needed to obtain the cross-section, and thus obtain an estimate of the effective size of the particles. We perform an experiment where an incident beam is scattered from a sample of known particle density n and thickness L . By measuring the fraction of incident particles which are transmitted through the sample, we can determine the collision cross section σ . For identical incident and sample particles $r_1 = r_2 = r$. A typical apparatus for this type of experiment is shown in figure 3.8.

3.3.1 Estimating atom size by other methods

Since methods such as that described above involve simplifications, it is worth having independent methods to verify the results. One possibility is to use the equation of state of a gas, but this time away from the approximations set by the ideal gas. One example for an equation of state describing a non-ideal gas is the Van der Waals equation

$$(P + \frac{a}{V^2})(V - b) = NkT \quad (3.17)$$

which contains two additional constants b and a . The quantity b represents the reduction of the volume of the gas by the volume of the gas particles themselves. It is proportional to the number of particles in the gas and to their volume. By varying the pressure, temperature and volume of a gas b can be determined experimentally, thus allowing the size of the particles in the gas to be obtained. Note that the constant a results in an increase of the pressure of the gas – it accounts for interactions (such as Van der Waals forces) between gas particles. As a and b tend to zero, equation 3.17 reduces to the equation of state for the ideal gas.

Another alternative, as alluded to earlier in a different context is to place a limit on the atomic size from the spacing of atoms in a crystal. Here, the maximum atomic size must be equal to the smallest lattice spacing which is obtained using Bragg scattering.

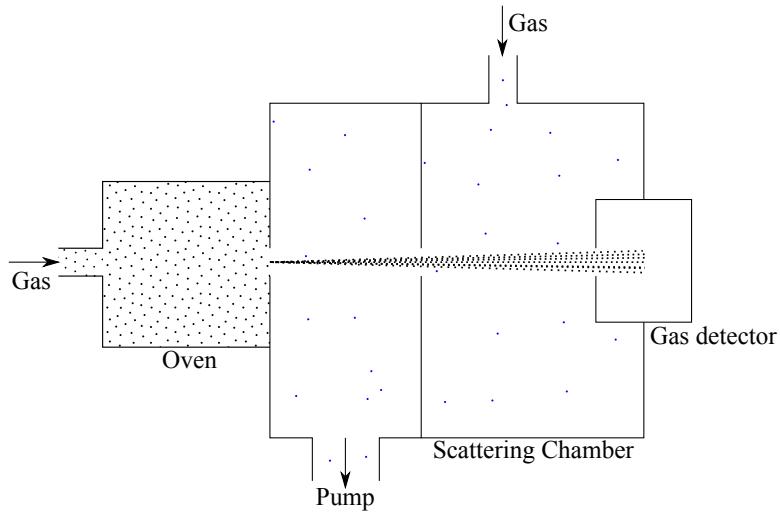


Figure 3.8: Setup with which the cross-section of gases can be probed. A beam of one kind of particles is generated by letting particles from a hot gas effuse from an oven. It passes a scattering chamber that contains another type gas. The number of atoms remaining in the beam is measured, e.g. by the pressure that it exerts on a surface.

3.3.2 Comment on the size of the atom

All of the aforementioned methods produce slightly different results, but the order of magnitude they give for the size of an atom is typically a few Angstroms (10^{-10} m). The reason that the measurements differ slightly is that they all test the “size” of the atom in a different way. This brings into question what we really mean by size. In the scattering experiments we see above, it is a region of influence, in that an object entering that region will be significantly deflected. The measured size differs because the interaction for an X-ray photon looks very different to that for a different atom.

You could also think about the electron – it is charged, so influences particles which are far away from itself. With the electron we know that so long as we are outside the area filled by the electronic charge density itself, we expect that other charges interact according to a $1/r$ potential. If we could get inside the electron, we would see deviations from this, but no-one ever has. It *is* worth thinking hard about what we mean by size on the atomic scale.

3.4 Evidence for atomic structure

The evidence for the structure of the atom was built up over a number of years in a large number of experiments, many more than we could possibly describe here. Some of these experiments you have seen already – the electrons which are observed in Compton scattering would indicate that electrons are constituents of atoms.

In general, if you want to find out if an object has internal structure, the way to do it is to look for inelastic scattering. If only elastic scattering is observed, then all the kinetic energy before the collision is converted to kinetic energy afterwards. If there is internal structure, some energy can be used up in re-configuring the internal components, and the scattering is then inelastic.

3.4.1 Electron scattering

Early electron scattering experiments were performed by Lenard, in the 1890s. He used a cathode-ray tube similar to that we discussed in the production of X-rays (in fact Lenard's work was the direct precursor to that of Röntgen). In Lenard's setup (shown in figure 3.9), a hole in the anode allowed electrons to pass through the electrode to a thin aluminium window (thickness 0.5×10^{-6} m) which formed the wall of the vacuum tube. The electrodes were not only able to pass through the foil window, but were observed to penetrate the air to a distance of a few centimeters (observed as a bluish glow close to the output of the tube). Using the cross-section obtained from atom scattering experiments, the electrons would be expected to penetrate the air to a distance of 10^{-7} m, which is 5 orders of magnitude smaller than what is observed. This indicates that the electron has a smaller cross-section for scattering from the atoms than is exhibited by atom-atom collisions.

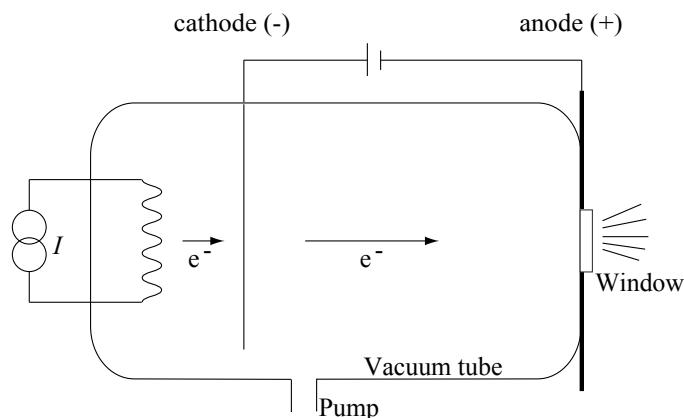


Figure 3.9: Sketch of a vacuum tube with a window (made out of a thin aluminium foil) after which electrons could still be observed.

More systematic experiments are able to control the pressure of the gas and the energy of the electrons, and look at the penetration as a function of both. The principal observations which are made are

- The cross-section for scattering is proportional to pressure. This agrees with the results for atom scattering. Pressure is either increased by increasing the number density of atoms, or by increasing the average speed of the individual atom. It is fairly obvious that the presence of a higher density of atoms increases the number of atoms which the electrons can scatter from. The dependence on atomic speed is more subtle, because

thus far we viewed the target atoms as being stationary. If they move, they are also sweeping out a certain volume, which leads to an increased chance of a collision.

- The cross-section is proportional to the density of the material or gas which is probed. The properties of the material, or the phase of matter do not make a difference.
- The cross-section is a function of the electron beam energy. It falls dramatically for higher energy electrons. As the speed of the electrons increases towards the speed of light, the cross-section falls by a factor of 10^8 . This drop is shown qualitatively in figure 3.10.

The final point is very important – it implies that only a small proportion of the volume of the atom is impenetrable for fast electrons. Note that previously the assumption had been that the electrons were smeared out over the whole atom – that matter was a kind of continuous medium. A possible classical explanation for the reduced scattering of high energy electrons is that the electrons are point particles which are moving around. This means that a faster electron would have a reduced chance of colliding, because it would spend less time in the area in which the electron moves. As we shall see later on, there were more surprises than this in store for experimentalists. I leave you with the impression of Lenard, who stated that “the inside of the atom is as empty as inter-planetary space”.

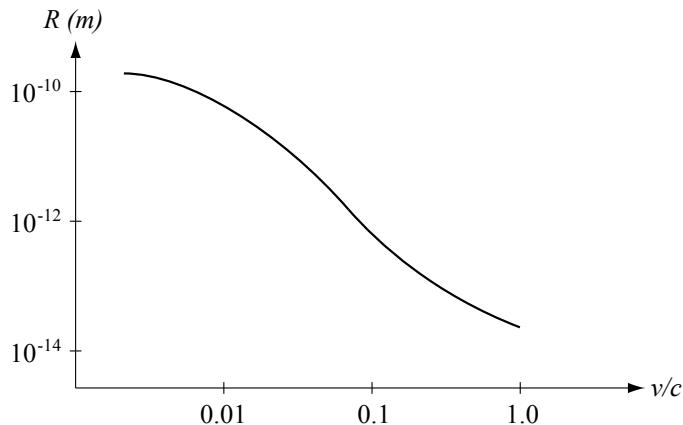


Figure 3.10: The scattering radius $R = r_1 + r_2$ for gas atoms and electrons as a function of the electron's velocity v . R is linked to the cross section by $\sigma = \pi R^2$.

3.4.2 The nucleus - Rutherford scattering

One of the most important experiments in determining the composition of atoms was performed by Ernest Rutherford in 1909. The experiment involved α -particles incident on a thin film of metal (thin being less than 10000 atoms thick). α -particles consist of 2 neutrons and 2 protons, and are thus positively charged. Rutherford was able to show that the angular deflection of the particles was consistent with most of the film being filled with empty space, with scattering occurring from localized centers of charge. We now know that this charge center is the nucleus of the atom. It is worth noting here that because α -particles are very heavy compared to electrons, individual collisions with electrons do not significantly deflect their path.

In order to understand Rutherford scattering, let us look at the expected angular distribution we would calculate for the alpha particles after being scattered by a charge of infinite mass. We will consider the alpha particles to be non-relativistic, since this is the case used in the experiments (here you should note that alpha particles are much heavier than electrons, so for a given energy they will be less relativistic). The scattering in this case is shown in figure

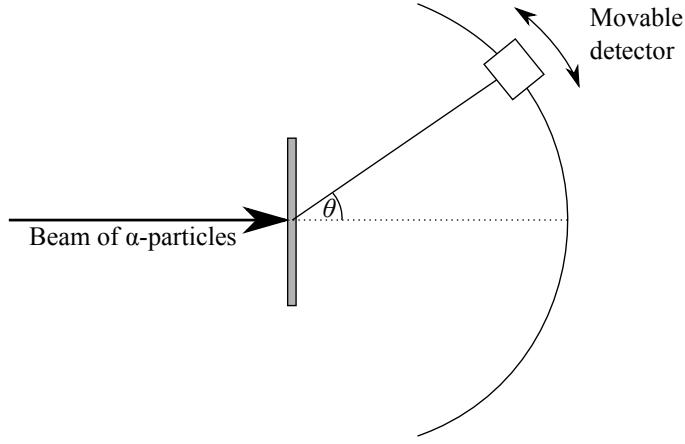


Figure 3.11: Schematic setup of Rutherford's scattering experiments. With a radioactive source a beam of α -particles is generated and scattered at a thin foil. The angular dependence of the flux of scattered particles is measured by a movable detector.

3.12. A charged particle with mass m and charge $2e$ is incident from the left of the figure. It has an initial speed v_0 , meaning that the initial angular momentum about the charge Ze is mv_0b , where b is the impact parameter shown in the figure (you should be familiar with these from your studies of classical mechanics). We will assume throughout that the charge Ze is fixed, as the atoms in a foil would be.

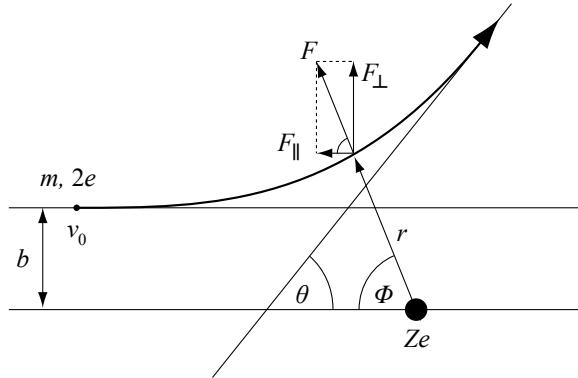


Figure 3.12: Schematic to show the geometrical quantities needed in the derivation of Rutherford's scattering formula.

The force on the alpha particle is given by

$$\mathbf{F} = \frac{2Ze^2}{4\pi\epsilon_0 r^2} \frac{\mathbf{r}}{r} = \frac{k \mathbf{r}}{r^2} \quad (3.18)$$

where \mathbf{r} is a vector from the charge Ze to the α -particle and r is the length of this vector. k is introduced as a constant to make the equations that follow easier to read. Since the force is radial, we know that angular momentum must be conserved. Therefore

$$m\dot{\phi}r^2 = mv_0b \quad (3.19)$$

which when re-arranged gives

$$\frac{1}{r^2} = \frac{\dot{\phi}}{v_0 b} \quad . \quad (3.20)$$

Now let us consider the force on the particle. We consider only one component of velocity, which is that perpendicular to the initial direction of travel. Calling this component v_{\perp} ,

we can quickly see from figure 3.12 that this component of the force is $|\mathbf{F}| \sin \phi$. Then from Newton's second law

$$\dot{v}_\perp = \frac{k}{mr^2} \sin(\phi) = \frac{k\dot{\phi}}{mv_0 b} \sin \phi \quad (3.21)$$

which gives

$$dv_\perp = \frac{kd\phi}{mv_0 b} \sin \phi . \quad (3.22)$$

If we are to integrate both sides of this expression, we should choose appropriate limits. For the left hand side, the initial value of v_\perp is zero, and if the particle starts at an infinite distance, the initial value of ϕ is zero. For a final angle of deflection of θ the final value for ϕ is $\pi - \theta$ (in the limit that the particle is at infinite distance). In order to get an upper limit for the integral over velocity we note that this is an elastic collision, and thus the total energy will be conserved. If the particle finishes infinitely far from the charge Ze all of the energy will be kinetic, and thus the velocity is again v_0 . Thus the upper limit on v_\perp is given by $v_0 \sin \theta$, and we find

$$\begin{aligned} \int_0^{v_0 \sin \theta} dv_\perp &= \int_0^{\pi - \theta} \frac{kd\phi}{mv_0 b} \sin \phi \\ v_0 \sin \theta &= \frac{k}{mv_0 b} (1 - \cos(\pi - \theta)) = \frac{k}{mv_0 b} (1 + \cos \theta) \\ b &= \frac{k}{mv_0^2} \cot(\theta/2) \end{aligned} \quad (3.23)$$

where in the last line I used a couple of trigonometric identities. We now have a definite link between the incoming impact parameter and the angle of scattering. If we wish to find which particles are scattered into a given angle between θ and $\theta + d\theta$ it is also useful to obtain the differential of the impact parameter with respect to θ , giving

$$db = -\frac{k}{2mv_0^2} \frac{1}{\sin^2(\theta/2)} d\theta . \quad (3.24)$$

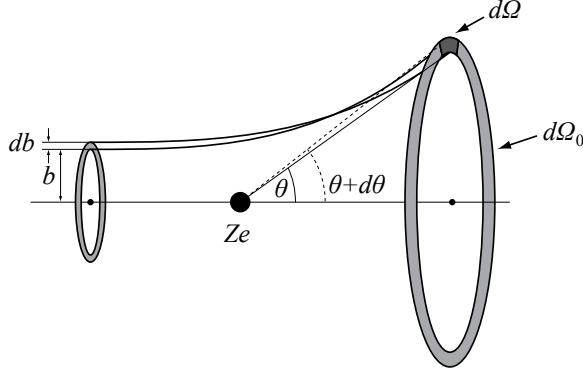


Figure 3.13: Illustration of how the impact parameter corresponds to an element of the solid angle.

So we've solved the scattering problem for two particles, but in a real experiment it is more typical to use an incoming beam of α -particles with a certain flux F , where this quantity is defined as the number of particles per second which pass through a unit area perpendicular to the beam. For a beam with N particles per second passing through an area A , the flux is $F = N/A$ and we find that the number per second with an impact parameter between b and $b + db$ with respect to a given atom in the sample is

$$dN_b = 2\pi F b db . \quad (3.25)$$

The particles with impact parameter b with respect to the atom will be scattered at an angle θ according to equation 3.23. Inserting equations 3.23 and 3.24 into the right hand side of this equation we see that the number of incoming particles per second scattered at angles between θ and $\theta + d\theta$ is

$$dN_\theta = dN_b = \pi F \left(\frac{k}{mv_0^2} \right)^2 \frac{\cos(\theta/2)}{\sin^3(\theta/2)} d\theta . \quad (3.26)$$

This is the number per second scattered into a total solid angle $d\Omega_\theta = 2\pi \sin \theta d\theta$. To find the number per second scattered into an element of solid angle $d\Omega = \sin \theta d\phi d\theta$ (here again figure 3.13 is useful), we use

$$\begin{aligned} dN_\Omega &= \frac{dN_\theta}{d\Omega_\theta} d\Omega \\ &= \pi F \left(\frac{k}{mv_0^2} \right)^2 \frac{\cos(\theta/2)}{\sin^3(\theta/2)} \frac{d\Omega}{2\pi \sin \theta} \\ &= \frac{F}{4} \left(\frac{k}{mv_0^2} \right)^2 \frac{1}{\sin^4(\theta/2)} d\Omega . \end{aligned} \quad (3.27)$$

Our original aim was to find the Rutherford scattering cross-section as a function of angle. Remember that the number scattered per second is given by the product of the flux and the cross-section, therefore we can obtain the cross-section from equation 3.27 by dividing both sides by the particle flux. This gives us an expression for the differential cross-section

$$d\sigma = \frac{1}{4} \left(\frac{k}{mv_0^2} \right)^2 \frac{1}{\sin^4(\theta/2)} d\Omega \quad (3.28)$$

If there are n_a atoms in the sample, and they do not overlap each other (the sample is “thin”) and a total of N in an incoming beam of area A , the number of particles scattered into solid angle between Ω and $\Omega + d\Omega$ is given by

$$N n_a \frac{d\sigma}{A} \quad (3.29)$$

You should be careful to note that the Rutherford formula is only valid if the interaction between the α -particles and the atoms is a Coulomb repulsion. In order to obtain a value for the nuclear size, we look to confirm the Rutherford formula for small values of the impact parameter b . This tells us that we still have $1/r$ type Coulomb law behavior, and therefore must be outside the charge distribution of the nucleus. As can be seen from equation 3.23, by using higher energies (though making sure these are still non-relativistic) we can constrain the maximum value of b for which non-Coulomb behavior could occur. Energies up to 5 MeV have now been used at angles out to 150 degrees, with no deviation from the Rutherford formula observed. This implies agreement with Coulomb forces for impact parameters of $b > 4 \times 10^{-15}$ m, and thus constrains the nuclear charge distribution to be within a volume with dimensions which are smaller than this.

3.4.3 Electronic structure

So the atom consists of a very small nucleus, surrounded by some electrons. The next question is where do the electrons sit and why is the atom stable? The primary methods by which information about the inner workings of the atom were deduced were optical: using the spectroscopy methods which we learnt about in the optics part of the course to split up different colours of light which atoms emit. A very important independent verification of the structure of atoms came from careful low-energy electron scattering experiments performed by James Franck and Gustav Ludwig Hertz.

Optical spectra

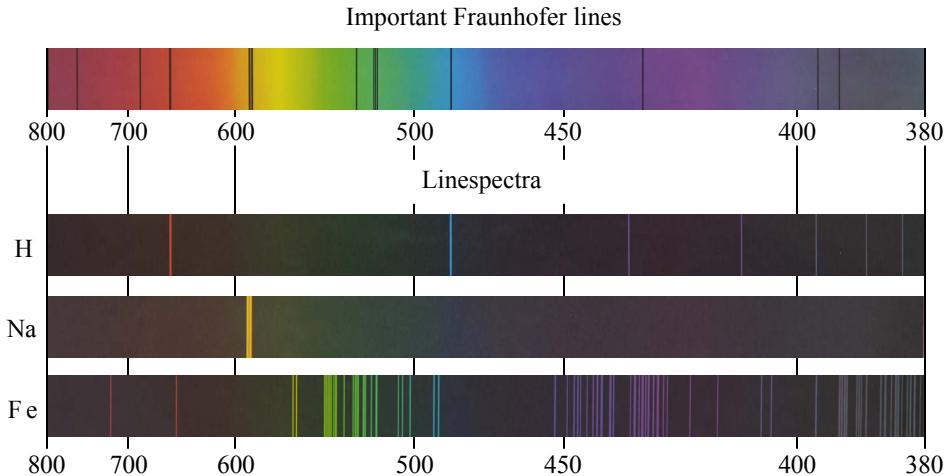


Figure 3.14: Top: Absorption lines in the sun's spectrum, named after Fraunhofer who carefully measured them. Below this are shown emission spectra of hydrogen, sodium and iron. Note that some of the lines in the solar spectrum coincide with the atomic emission lines. It is also worth noting that the complexity of the spectrum increases with atomic weight. This is because heavier atoms have many more electrons, and the complexity is thus increased.

Investigations of the electronic structure of atoms using light takes two forms. Either you force atoms to emit light, and measure the spectral components of this emission, or you take a known light source and look at how much of the light at different frequencies is absorbed by the atoms. From looking at these two types of spectra, researchers were able to deduce that atoms emit at a set of discrete frequencies. The set of frequencies is unique for each type of atom. This makes it possible to know what elements are contained in the atmosphere of our sun. The sun emits at a broad range of wavelengths, but some of these are absorbed in its atmosphere. By comparing absorption lines in the light from the sun to the spectral lines of elements that we know from measurements performed on earth, we deduce the atmospheric elements.

Figure 3.14 shows both the solar spectrum and emission spectra for hydrogen (H), sodium (Na) and iron (Fe). First note that the emission spectra of all three consist of well-spaced discrete lines. Secondly, note that many of the hydrogen emission lines are close to the absorption lines from the sun's light. This implies that there is plenty of hydrogen in the sun's atmosphere. Later in the course we will use quantum mechanics to predict the emission spectra of the hydrogen atom - the correspondence between theory and experiment in this example has been used over the past century to test quantum mechanics to an extremely high level of precision.

Franck-Hertz experiment

Independent evidence that the electronic structure of atoms consisted of discrete energy levels came from an experiment conducted by James Franck and Gustav Ludwig Hertz. Their experiment involved the passage of electrons through a “vacuum” tube filled with mercury vapor. The setup is shown in figure 3.15. The electrons are first emitted from E by thermal emission (E is heated up). Due to the potential difference between E and the grid G , these electrons gain kinetic energy eU on arrival at G . Electrons passing through the grid then lose kinetic energy eV before arriving at the collector plate C . If electrons are able to pass from E to C , current will flow, which is detected on the ammeter I . Of

course the description given ignores the Mercury atoms. Since these are much heavier than the electrons, they do not gain much energy from an elastic collision with an electron. This means that elastic collisions do not remove much kinetic energy from the electrons, and the number of electrons reaching the collector is unaffected, as is the measured current.

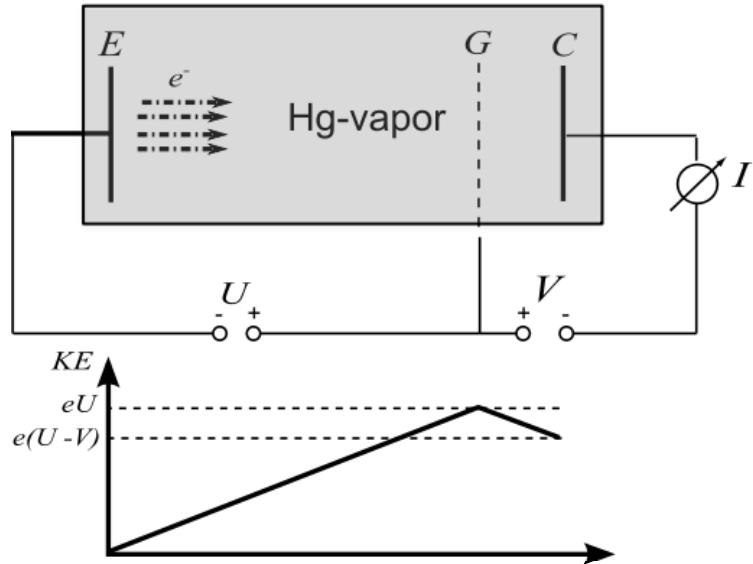


Figure 3.15: Setup and kinetic energy (KE) as a function of position in the Franck-Hertz experiment. Electrons from an emitter E are accelerated due to the potential difference U between the grid G and E . The electrons passing through the grid must climb a potential hill of eV before hitting the collector C . The electrodes are placed inside a vacuum chamber which is filled with a low density of Mercury vapor.

By contrast, since mercury atoms also have structure, it is possible for the electron to make an inelastic collision with the mercury atom. In this case, the mercury is reconfigured by the collision, changing its internal energy by an amount E_I . In this case the electron will lose energy E_I in the collision. If this energy loss is greater than $e(U - V)$, then electrons will no longer have the energy required to reach the collector, and the measured current will drop.

The measured current observed as a function of applied voltage U for the experiments of Franck and Hertz is plotted in figure 3.16. As U is increased from zero, we see that electrons start to reach C when $U > V = 0.5$ V. The number of electrons then increases as the voltage increases, up to a value of $U = 4.86$ V. At this point the current suddenly drops. The reason for this is that electrons in the region close to G now have an energy of 4.86 eV, which corresponds to the energy difference between two internal states of the mercury atoms. Thus inelastic collisions occur, and the electron energy is lost to the mercury, meaning that electrons no longer have the kinetic energy required to reach C .

As the voltage is increased beyond $U > V + E_I = 5.36$ V the current again increases, because even after an inelastic collision the electrons have enough kinetic energy to reach the collector. The next drop in current occurs when $U = 2E_I$. At this point electrons which make a first inelastic collision at the point at which they have gained 4.86 eV of kinetic energy subsequently gain an additional 4.86 eV of kinetic energy which means that they can make a further inelastic collision. As U increases further this cycle repeats periodically at each multiple of E_I .

In addition to inelastic collisions at values of U which are multiples of 4.86 eV, other drops

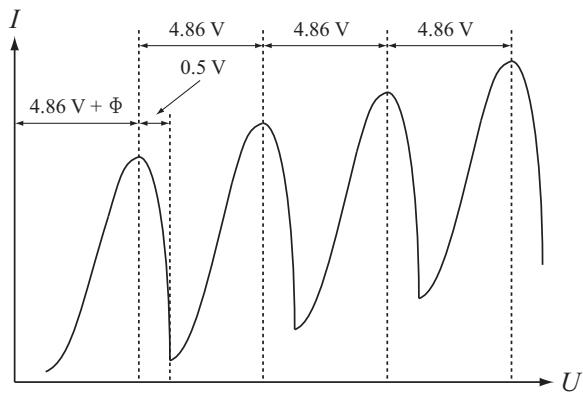


Figure 3.16: Dependence of the current I on the accelerating voltage U in the Franck-Hertz experiment. The dependence of the current on the voltage is described in the main text.

in current are observed at higher values which are not multiples. These correspond to other ways in which the internal structure of the mercury atoms can be re-arranged by the collision. Franck and Hertz were also able to measure UV light emitted from the mercury atoms at 4.86 eV, which corresponds to a wavelength of 255 nm.

Chapter 4

Statistical Mechanics

4.1 Books and problems

The best book which I have come across on statistical mechanics is “Statistical Physics” by F. Mandl. The lectures will follow the approach and derivations in that book, so I’d highly recommend getting hold of a copy. I’ve been told that a book by A. M. Guenault called “Statistical Physics” is at a better introductory level, so you may prefer this, especially as here we are trying to take an introductory approach to the subject, rather than an extremely rigorous one (which you will get elsewhere in your course).

In the lectures, we shall cover the concepts, but as ever, to learn the topic you will have to apply these concepts to problems. In my experience, this isn’t so easy, so it’s worth having books which work you through examples. We don’t have enough time to do this extensively in the lectures. The books mentioned above have many nice examples. It is tempting to just “do the maths” when working through examples, and this is the impression you get when reading many books. I warn you to be careful here. The difficulty in learning about statistical mechanics is that differences such as “isolated” and “in thermal equilibrium with a heat bath” are important, so you have to think hard about exactly what the assumptions are when doing a particular problem, or you will get some wrong answers.

4.2 Equilibrium statistical mechanics

Statistical mechanics is concerned with predicting the behavior of macroscopic systems from the microscopic details. Macroscopic parameters which we might have an interest in are pressure P , mass (or number of components N), the energy E or the volume V . In addition, thermodynamics uses quantities such as temperature T and entropy S which are important for the consideration of macroscopic systems. How should we go about predicting these values from the microscopic properties of collections of atoms and molecules?

For example, in a gas, one might consider taking each atom, and attaching a label i to it. Then we can specify positions and momenta of each atom p_i, q_i . Newton’s laws tell us that given all these parameters, we can predict the future motions of all the atoms. The need for a statistical approach arises because the number of atoms is typically greater than 10^{23} . It is clearly not possible to measure this number of positions and momenta, and even if we could then computing the future evolution would defeat any computer that will ever be built.

There are several important concepts in statistical mechanics. The first is the concept of a state, which can be split into two parts

- A *macrostate* is the macroscopic state, which is defined only by bulk variables, such as P, N, E, V .
- A *microstate* corresponds to a particular microscopic configuration. For atoms in a gas, this is specified by a list of the positions and momenta of all the atoms.

For a system of 10^{23} atoms, there are clearly a very large number of microstates. What is more, many of these will share all of P, N, E, V . So a single macrostate is consistent with a large number of microstates.

4.2.1 Coin example

To illustrate the concept of macro and microstates, think about a system of four coins, which can each have either the heads or tails state when laid flat on a table. Here the macrostate is defined by the number of coins N , which is fixed at four, and the number of the coins which are tails n (which lies between 0 and 4). There are 5 macrostates of the four coins available, each of which are fully specified by the macroscopic variable pair N, n . An example of a microstate is HHHT, which tells us that the fourth coin is tails but the other three are heads; note the microstate specifies *which* coin is tails. For each macrostate, there are many microstates. For example, the macrostate of the four coins given by $N = 4, n = 3$ has four microstates: HTTT, THTT, TTHT, TTTH.

The basis of statistical mechanics is the question “How many microstates correspond to each macrostate of the system” and applying a set of rules which attach some physical significance to the answer. In the hands of Boltzmann, Gibbs and others this approach was phenomenally successful in accounting for a wide range of phenomena.

Before pursuing statistical mechanics in more depth it is worth reminding ourselves about probability distributions. In addition, though I’m talking about counting discrete numbers of microstates here, some of you may have spotted already that position, momentum, energy and volume are continuous variables. This is something we should clarify. Once we have done this, we shall look at an example problem of relating microscopic to macroscopic variables in an ideal classical gas, and see what contributions are required from statistical mechanics in order to aid our understanding of this situation.

4.2.2 Continuous and discrete probability distributions

Before we get stuck into the details of statistical mechanics, it is worth reminding ourselves about the tools. One of the principal tools is the probability distribution. Think about a dice. Here we say that there is an equal probability of getting each of the six numbers on the dice. Therefore the probability distribution looks like

$$p_n = \frac{1}{6} = \frac{1}{\text{total number of options}} \quad (4.1)$$

for any n from 1 to 6. If I want to find what the average value of the number that I will roll is, I take

$$\langle n \rangle = \sum_n np_n = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5 \quad (4.2)$$

Now let us move to a continuous problem from physics. For a particle in free space about which we have no prior knowledge, we might want to ask “with what probability is the particle at position x ?” These are exactly the sort of questions we ask in statistical mechanics. However it is not a clever question, because the answer can be specified to arbitrary precision, and thus the probability is infinitesimal. Also it makes no sense to ask this question. In physics, we are interested in what we can measure physically, and we can never measure to arbitrary precision.

The sensible question is instead to ask “what is the probability that x lies between x and $x + dx$?” The probability is given by

$$p(x)dx \quad (4.3)$$

where $p(x)$ is a continuous probability distribution. We know that probabilities must sum to equal 1, ie.

$$\int_{-\infty}^{\infty} p(x)dx = 1 \quad (4.4)$$

The average value of x for the particle is then given by

$$\langle x \rangle = \int_{-\infty}^{\infty} xp(x)dx \quad (4.5)$$

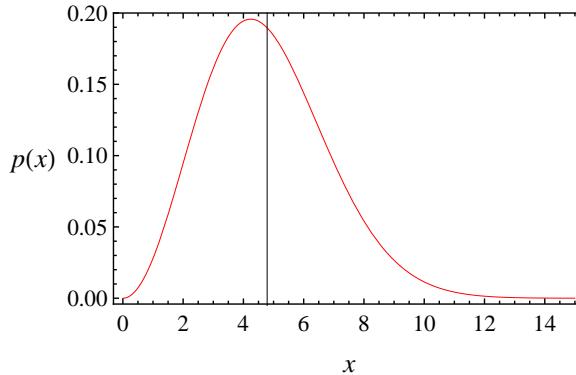


Figure 4.1: An example of a continuous probability distribution. A relevant question to ask is “what is the probability to find an object between x and $x + dx$, which is given by $p(x)dx$. The mean is marked by a vertical line. Note that this not the most probable value.

4.3 Kinetic theory and the ideal gas

In your thermodynamics course last year, you saw experimentally that for a simple gas, several macroscopic quantities are related by the *equation of state*

$$PV = Nk_B T \quad (4.6)$$

where N is the number of particles in the gas, P is the pressure, V is the volume, and T is the temperature (quoted on the Kelvin scale). Here, since we are considering links between microscopic and macroscopic, we will try to derive this formula from the consideration of the motion of the particles themselves. We start by considering a single particle in a box. It has velocity $\mathbf{v} = (v_x, v_y, v_z)$ and mass m . Let us consider the *average* force exerted by this particle on the right hand wall of the box, which is shown in figure 4.2. When the atom

bounces off this wall of the box (elastically), it gets a change of momentum in the horizontal z direction

$$\delta p_z = 2mv_z . \quad (4.7)$$

From Newton's laws, we know that the force on the wall of the box is equal and opposite to the force on the particle

$$F_z(\text{wall}) = -F_z = -\frac{\delta p_z}{\delta t} \quad (4.8)$$

so it is clear we need to obtain the time interval between collisions. A particle with this velocity in z hits the same surface again after it has crossed the box to get to the other side and come back again. This will take a time

$$\delta t = \frac{2L}{v_z} \quad (4.9)$$

and therefore the force on the right hand wall of the box, averaged over time is

$$F_z(\text{wall}) = -\frac{mv_z^2}{L} . \quad (4.10)$$

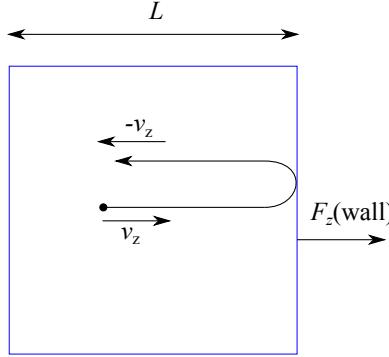


Figure 4.2: A particle in a box with sides of length L changes its momentum by $2mv_z$ when hitting a wall and therefore exerts a force F_z onto the wall.

Now consider that instead of knowing that the particle has z velocity v_z , we don't know the velocity, but know it follows a *probability distribution* $p(v_z)$ such that the probability to be found between the values v_z and $v_z + dv_z$ is $p(v_z)dv_z$. We can then find the expectation value of the force F_z for the particle

$$\langle F_z(\text{wall}) \rangle = -\frac{\int_{-\infty}^{\infty} mv_z^2 p(v_z) dv_z}{L} . \quad (4.11)$$

The pressure is given by $F_z(\text{wall})/A$, resulting in

$$P_z = \frac{\int_0^{\infty} mv_z^2 p(v_z) dv_z}{V} = \frac{m}{V} \langle v_z^2 \rangle \quad (4.12)$$

Note that $m\langle v_z^2 \rangle / 2$ is the kinetic energy for the z component of particle motion.

It is very rare for a gas to consist of a single particle. For a gas containing N similar but *distinguishable* particles, and *each having the same probability distribution*, we would find

$$P_z = \frac{Nm}{V} \langle v_z^2 \rangle \quad (4.13)$$

You can see from this treatment that it is important to be able to extract probability distributions, because they can tell us about bulk properties. Now we need to figure out

how to find these for systems of relevant interest. At this stage in the course, we shall consider a gas containing a large number of particles in *thermal equilibrium* at temperature T . We shall see that this allows us to find that

$$\frac{1}{2}m\langle v_z^2 \rangle = \frac{k_B T}{2} \quad (4.14)$$

where k_B is the Boltzmann constant. This leads us to

$$PV = Nk_B T \quad (4.15)$$

which agrees with equation 4.6. In the following sections we will fill in the gaps. We will determine the probability distribution, and use it to derive equation 4.14. Thus we will be able to arrive at the equation of state for the ideal gas from purely microscopic considerations.

There are several highlights in the previous sections, concerning probability distributions, distinguishability and thermal equilibrium. We should be very careful when reading texts - statements such of these are critical to the physics, but if you read sentence they are not highlighted they can easily be missed.

4.4 The Boltzmann distribution

The Boltzmann distribution gives the probability to find the system in a micro state with a given energy for a system in thermal equilibrium. In order to derive this distribution, we start from the following statement, which is probably the most important in statistical mechanics. Since it is so important, I put it in quote marks AND italics.

- “*For a closed system, which has fixed constraints (e.g. fixed energy, particle number, volume), each microstate which satisfies the constraints has an equal probability of being occupied.*”

The question in statistical mechanics is what implications this has for the properties of the system. One example might be to find out the probability that the system is in a particular *macrostate*. The next important statement from statistical mechanics (which we shall see an example of in the next section) is that

- “*The equilibrium macrostate of a system will be the macrostate which satisfies the fixed constraints and has the maximum number of microstates.*”

The implication is that all we need to do to find the equilibrium macrostate of the system is to count the number of microstates which are consistent with its parameters. This statement is an extremely good approximation when the number of microstates available gets very large.

4.4.1 Coin flipping and the equilibrium state.

In order to illustrate the thinking involved, let us go back to the simple example of our four coins, each of which can either be heads or tails (edges not allowed). We set them in a given configuration, and then arrange a game such that on every time step we choose one of the four coins at random and flip it. This is completely reversible (each coin has an equal chance of being flipped $H \rightarrow T$ if it is already H , and vice versa).

We now ask which is the most likely macrostate to end up in in the steady, or equilibrium state? The steady state is the one for which the probabilities to be in any particular macrostate are not changing with time. In this example there are a couple of parameters which define each macrostate. One is the number of coins N , which is fixed at four. Another parameter which defines different macrostates but is not fixed is the number of coins which are tails n . For a macrostate with two tails, we have six microstates, namely HHTT, HTHT, HTTH, THHT, TTHH, THTH. The number of microstates for each macrostate is given in table 4.1.

Macrostate, no. of tails	0	1	2	3	4
No. of microstates	1	4	6	4	1

Table 4.1: Number of microstates for each macrostate of the 4-coin experiment.

Let us form an equation which describes the jump in occupation probability at a single step for each of the macrostates. We will denote the probability of being in the macrostate with n tails by $P(n)$. For $P(4)$, we have that the probability of a transition from $n = 3$ is $P(4|3) = 0.125$, since it requires the only head in the $n = 3$ macrostate to flip. Going the other way, we find that $P(3|4) = 0.5$, because whichever of the coins we choose to flip

to head will result in our ending up in one of the states of $n = 3$. The equation for the evolution of $P(4)$ for a single time step is therefore

$$\Delta P_{t+1}(4) = P(4|3)P_t(3) - P(3|4)P_t(4) \quad (4.16)$$

therefore in the steady state (when the probabilities are independent of time, and $\Delta P_\infty(4) = 0$

$$P_\infty(4) = \frac{P(4|3)P_\infty(3)}{P(3|4)} = \frac{0.125}{0.5} P_\infty(3) = 0.25 P_\infty(3) \quad (4.17)$$

where I have discarded the time subscripts because in the steady state these don't matter.

A similar analysis performed on the $P(3)$ macrostate gives

$$P_\infty(3) = \frac{2}{3} P_\infty(2) \quad (4.18)$$

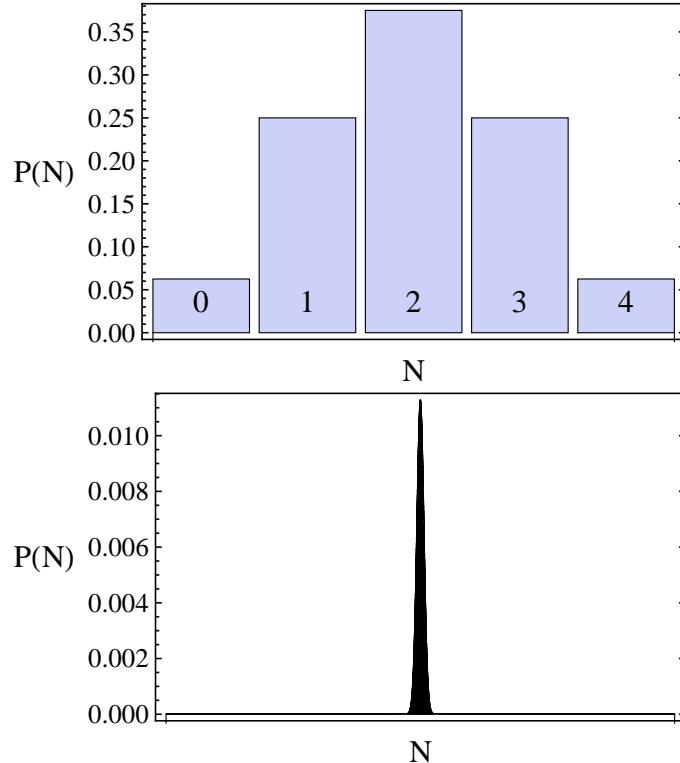


Figure 4.3: Steady state probabilities for the coin-flipping game. a) For 4 coins, we see a binomial distribution centered at 2, which is the macrostate with the highest number of microstates. b) For 5000 coins, we see that the distribution is centered at a macrostate with $N = 2500$, but the fractional width $\Delta N/N$ is now much smaller. For 10^{23} coins, the peak would be extremely sharp (I cannot compute this though, so you don't get a plot).

By symmetry, we can also see that $P_\infty(1) = (2/3)P_\infty(2)$ and $P_\infty(0) = (1/4)P_\infty(1)$, therefore the most probable macrostate has $n = 2$. In this case, the probability distribution for the steady (equilibrium) state is shown in figure 4.3. Also shown is the result of the same experiment for 5000 coins. Note that the ratio of the width of the peak to the mean value becomes much smaller as N is increased - this means that the macrostates of importance are the ones close to the most likely one to be occupied, which brings us back to the original postulate. This becomes a better approximation as the numbers increase, and is extremely

good for $N \simeq 10^{23}$ (the fractional width of the peak (width divided by center value) can be shown to scale as $N^{-1/2}$). Note that once we state the total number of coins, the number of heads is to a good approximation a completely specified parameter *as long as we have attained thermal equilibrium*.

4.4.2 An isolated system of distinguishable particles

In this section we will consider an isolated system made up of a large number of distinguishable objects. This could for instance describe atoms in a solid, because the locations of each are separate (we can therefore always identify which atom we are talking about). Nevertheless each atom can move, and might have energy associated with being a certain distance from its equilibrium position (for example if it is bound in a harmonic oscillator potential). Therefore we have a number of different position/momentum pairs that we can find it at which define the atom's energy. Our aim is to try to find the probability of finding each atom at a given energy conditioned on the atoms being in thermal equilibrium with each other. Since there are so many atoms in the solid, we cannot go in and measure the exact energy of each atom. For simplicity, we shall assume that the number of atoms is fixed, and the total energy of all the atoms is also fixed.

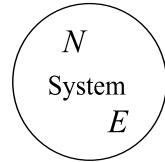


Figure 4.4: A closed system that has a fixed number of particles N and a fixed total energy E .

A microstate of the system might look like n_1 atoms described by the same position/momentum pair variables corresponding to the state numbered 1, n_2 with pair corresponding to state 2, up to n_K in state K . If we sum up the n_i , we get a total number of atoms $N = \sum_i^K n_i$. The total energy of this configuration would be $E = \sum_i^K e_i n_i$, where e_i is the energy of an atom in state i . The macrostate of this system is defined by the fixed total number of atoms, the fixed total energy, and the list of numbers n_i . Note that this does *not* define a microstate because we didn't say *which* atoms were in 1, 2, 3... etc..

Let us now try to count the number of microstates for a given macrostate. Suppose you start with the N objects, and ask how many ways can I distribute these in a set of K buckets such that n_1 are in bucket 1 etc.. You'd start by picking out n_1 from the N objects, and put them in bucket 1. How many ways can you do this? The answer comes from probability you probably did at school. If $n_1 = 1$, there are N ways to do this. If $n_1 = 2$, when picking the first object to go in bucket 1 there are N options, but when you come to pick the second there are only $N - 1$ remaining, hence the number of ways seems to be $N(N - 1) = N!/(N - 2)!$. However you also note that it doesn't matter which object got picked first, the final state is that a particular two objects are in the box. Therefore the real number is $N!/(2!(N - 2)!)$. After making the choice for the first n_1 objects, you have to choose n_2 objects from those remaining, of which there are now $N - n_1$. The total number of ways to assign all the objects to all the boxes is then

$$\begin{aligned} \Omega &= \frac{N!}{n_1!(N - n_1)!} \frac{(N - n_1)!}{n_2!(N - n_1 - n_2)!} \frac{(N - n_1 - n_2)!}{n_3!(N - n_1 - n_2 - n_3)!} \cdots \frac{n_K!}{n_K!0!} \\ &= \frac{N!}{n_1!n_2!n_3!\dots n_K!} \end{aligned} \quad (4.19)$$

We will now try to maximize this number with respect to the occupancies n_i . We will find

it convenient to work with the logarithm of this quantity, which is

$$\ln \Omega = \ln(N!) - \sum_i \ln(n_i!)$$
 (4.20)

We are particular about the solution we are looking for, in the sense that we wanted to know the probability distribution under the constraint that the energy is E and the number of atoms is N . We therefore need to look for a maximum of $\ln \Omega$ which simultaneously satisfies these constraints. To do this we use the method of Lagrange multipliers. The maximization therefore consists of looking for the solution to the set of equations

$$\begin{aligned} 0 &= \frac{\partial}{\partial n_i} \left(\ln \Omega + \alpha(N - \sum_i n_i) + \beta(E - \sum_i e_i n_i) \right) \\ &= \frac{\partial \ln(n_i!)}{\partial n_i} + \alpha + \beta e_i \end{aligned}$$
 (4.21)

You are probably not very familiar with taking the derivative of the logarithm of the factorial of a number. It turns out for a situation where n_i is large, there is a relation known as Stirling's theorem which helps us out. It looks like

$$\ln(x!) \simeq x \ln x - x .$$
 (4.22)

Using this formula, but noting that it does require a large number to be valid, we find that the set of equations we are required to solve is

$$0 = -\ln(n_i) - \alpha - \beta e_i$$
 (4.23)

which when solved for n_i yields

$$n_i = \exp(-\alpha - \beta e_i)$$
 (4.24)

So the equilibrium state is such that the number of atoms in state i with energy e_i is proportional to the negative exponent of the energy of the state. If N is large, then the numbers resemble the probability distribution, and we can therefore obtain a probability that a given atom is in the i th state

$$p_i = \frac{n_i}{N} = \frac{\exp(-\alpha - \beta e_i)}{\sum_i \exp(-\alpha - \beta e_i)} = \frac{\exp(-\beta e_i)}{Z_{SP}}$$
 (4.25)

where

$$Z_{SP} = \sum_i \exp(-\beta e_i)$$
 (4.26)

is known as the *single particle partition function*. The sum is taken over all the possible states that the atom can be found in.

In order to make any predictions, it is clear we need to find the value of the constant β . To do this, we will expand the problem as we would in thermodynamics, by adding a “bath” system with which the system of interest can exchange energy but which remains at a constant temperature T .

4.4.3 A system in thermal equilibrium with a heat bath

Let us imagine we have one “system” embedded within another, much larger “bath” system. The combination of both are completely isolated from anything else. The total energy of system plus bath is E . The small system, though smaller than the bath, will still have a large number of particles, so there is no way we could possibly track its dynamics. The

small system will have a fixed volume and particle number, which are well known. Let us first find the probability for this system to be in a given microstate, which we will label with s and take to have energy E_s

To do this, consider that the small system does have energy E_s . This means that the bath has energy $E - E_s$. Since there are many particles in the bath, there are a large number of microstates of the bath which will have this energy. We shall call the number of these states $\Omega(E - E_s)$. If all states of the complete system are equally likely to occur, then the probability to find the system at energy E_s is proportional to the product of the number of corresponding states in the bath and the number of corresponding states of the system.

$$p_s = A\Omega(E - E_s) \quad (4.27)$$

where A is a constant. In order to find A , we note that probabilities should sum to 1, ie.

$$p_s = \frac{\Omega(E - E_s)}{\sum_s \Omega(E - E_s)} \quad (4.28)$$

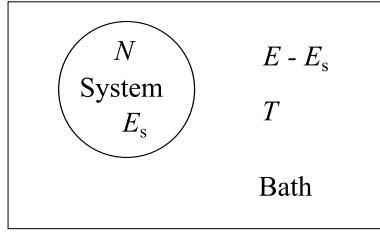


Figure 4.5: The system (circle) is embedded in a large bath, with which it can exchange energy. The bath is so big that the exchange of energy with the system does not change its temperature T . It is important to notice that the closed system now consists of the system plus bath.

The next step is a postulate, which is written on Boltzmann's gravestone in Vienna. He defined a relation between the number of states and the entropy

$$S(E - E_s) = k_B \ln \Omega \quad (4.29)$$

This is probably the most famous equation in statistical mechanics. It is important to remember that it is a postulate, he did not derive it. However it has come to be generally accepted due to the fact that the predictions made from this relation match the experimental evidence extremely well.

On combining equations 4.29 and 4.27, we find that the probability of occupation of the s microstate is

$$p_s = Ae^{S(E-E_s)/k_B} . \quad (4.30)$$

The discussion up until now doesn't have any assumptions about the size of the "heat bath", or the "system". Now we will make an assumption, which will show the reason why they are so named. The two systems do exchange energy continuously, but we will assume that though this changes the energy of the system by a significant fraction, it has little effect on the bath. The explicit assumption is that $E_s \ll E$. It is obvious that if energy can be exchanged then there must be some states for which this condition is not valid (for instance when all the energy is in the system, and none in the bath). However we work under the assumption that these are extremely unlikely to occur, and this is certainly the case when system and bath are in thermal equilibrium. This assumption allows us to expand the entropy around E

$$S(E - E_s) \simeq S(E) - \frac{\partial S(E)}{\partial E} E_s + \frac{\partial^2 S(E)}{\partial E^2} \frac{E_s^2}{2} . \quad (4.31)$$

From thermodynamics, we know that for a bulk system

$$\frac{\partial S(E)}{\partial E} = \frac{1}{T} . \quad (4.32)$$

If the bath stays in thermal equilibrium, its temperature does not change as a result of interactions with the system. Therefore, the second derivative

$$\frac{\partial^2 S(E)}{\partial E^2} = \frac{\partial}{\partial E} \left(\frac{1}{T} \right) = 0 . \quad (4.33)$$

The net result of this is that the expansion of the entropy looks like

$$S(E - E_s) \simeq S(E) - \frac{E_s}{T} \quad (4.34)$$

and on substituting back into equation 4.30

$$p_s = A e^{S(E)/k_B - E_s/(k_B T)} = A' e^{-E_s/(k_B T)} . \quad (4.35)$$

Now we should stop, and look back at equation 4.25. The form of the equation is very similar to this one, so long as we make the identification that $\beta = 1/k_B T$. This gives us the crucial missing piece in the jigsaw which we were lacking from the previous section. Note that in the case here, we only needed to assume that the bath was a large system, whereas due to Stirling's theorem in the previous section, we had to worry about the system size.

Making use again of the fact that probabilities must add to equal 1, we find

$$p_s = \frac{e^{-E_s/(k_B T)}}{\sum_s e^{-E_s/(k_B T)}} . \quad (4.36)$$

This is the Boltzmann distribution, it is incredibly famous, and you should remember it until you die. The factor in the denominator is known as the partition function of the system and is often written as

$$Z = \sum_s e^{-E_s/(k_B T)} . \quad (4.37)$$

Let me emphasize the meaning of equation 4.36: the probability here is the probability that the system is in the microstate s given that it is in equilibrium with the heat bath. This is different to the probability that we found in the previous section, which was for a given atom to be in a given state. However, if we think of the atom as our (very small) system in thermodynamic equilibrium with all the other atoms (which serve as a bath) we could identify its state with the microstate here.

If we were to consider the probability to find the system in any of a number $\Omega_s(E_s)$ of microstates with energy E_s , then the probability of being in one of these states is

$$p_{E_s} \propto \Omega_s(E_s) \Omega(E - E_s) . \quad (4.38)$$

We can follow through the same working from equations 4.29 to 4.36, only this time we find that

$$p_{E_s} = \frac{\Omega_s(E_s) e^{-E_s/(k_B T)}}{Z} \quad (4.39)$$

where the new form for the denominator is

$$Z = \sum_{E_s} \Omega_s(E_s) e^{-E_s/(k_B T)} . \quad (4.40)$$

Note the difference between this equation and the previous definition of Z , in equation 4.37. Here we have the extra factor $\Omega_s(E_s)$, which is accounted for by the fact that we sum over the energy, not the individual states.

The Boltzmann distribution is shown in figure 4.6 for the case of zero degeneracy. Note that when the temperature is lower, the lower energy states are more likely to be occupied.

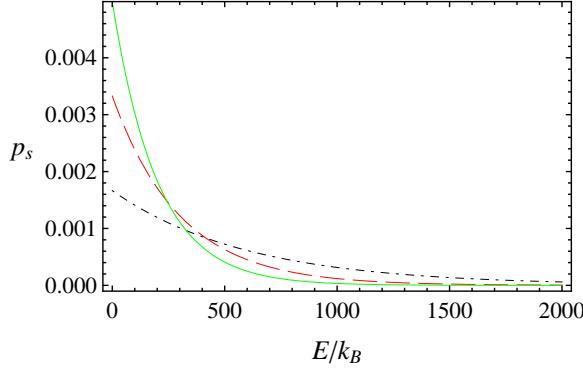


Figure 4.6: Boltzmann distribution for an equally spaced ladder of energy states with no degeneracy plotted for three temperatures. In order of decreasing temperature, the curves are (dot-dashed black), (dashed red) and (solid green). Note that for lower temperatures the low energy states are more likely to be populated.

4.4.4 Example of use of the partition function

The partition function is the starting point for deriving any macroscopic parameter of the system. For example, in order to find the average energy of the system, we can take

$$\langle E_s \rangle = -\frac{\partial(\ln Z)}{\partial \beta} . \quad (4.41)$$

This can be quickly seen from the form above, for which

$$-\frac{\partial(\ln Z)}{\partial \beta} = \frac{\sum_{E_s} \Omega_s(E_s) E_s e^{-E_s/(k_B T)}}{\sum_{E_s} \Omega_s(E_s) e^{-E_s/(k_B T)}} = \frac{\sum_{E_s} E_s p_{E_s}}{\sum_{E_s} p_{E_s}} \quad (4.42)$$

which is the mean energy, according to our previous knowledge of probability distributions.

4.5 The equipartition theorem

One of the most important results in classical statistical mechanics is the emergence of the equipartition theorem, and thus we follow through the derivation here. We consider a set of almost independent particles, in that we will ignore their interactions, but they should have interacted to some extent to ensure that they arrive at thermal equilibrium. The phase space co-ordinates of particle i are given by p_i, q_i , which refer to the momentum and position of the particle, as before. Let us assume that each particle is in a 1-D harmonic-oscillator potential so that it has energy

$$E(p_i, q_i) = \frac{p_i^2}{2m} + \frac{kq_i^2}{2} = ap_i^2 + bq_i^2 . \quad (4.43)$$

Now we need to consider what we mean by a state in this context. Remember that it is meaningless to specify the probability that the particle i is at position q_i with momentum p_i , but rather we specify probabilities to be between q_i and $q_i + dq_i$ with a momentum between p_i and $p_i + dp_i$. If this region is what we class as our “state” in continuous phase space, then the probability to be in this state is $p(q_i, p_i)dq_idp_i$. We know from the previous section that if our system is in thermal equilibrium, this probability follows a Boltzmann distribution, therefore

$$p(q_i, p_i)dq_idp_i = \frac{1}{Z_i} e^{-\beta E(q_i, p_i)} dq_idp_i \quad (4.44)$$

where Z is now

$$Z_i = \int \int e^{-\beta E(q_i, p_i)} dp_i dq_i = \int e^{-\beta ap_i^2} dp_i \int e^{-\beta bq_i^2} dq_i \quad (4.45)$$

and each integral is taken from $-\infty$ to ∞ . Note here that the integrals separate out, because the energy depends on the sum of energy terms due to the different co-ordinates.

Next we should think about looking at a collection of such particles. The energy of a many particle microstate is

$$E_s = \sum_i (ap_i^2 + bq_i^2) \quad (4.46)$$

where the sum over i sums over all the particles.

First let us consider the case that all the particles are distinguishable, but all have the same mass and spring strength (they might each be attached to a different spring which does not allow them to become tangled). In this case we can simply form the partition function for the many-particle system from the single particle system as

$$\begin{aligned} Z &= \int \int \int \dots \int e^{-\beta \sum_{i=1}^N ap_i^2 + bq_i^2} (dp_1 dq_1)(dp_2 dq_2) \dots (dp_N dq_N) \\ &= \int e^{-\beta ap_1^2} dp_1 \int e^{-\beta bq_1^2} dq_1 \dots \int e^{-\beta ap_N^2} dp_N \int e^{-\beta bq_N^2} dq_N \\ &= \left(\int e^{-\beta ap^2} dp \int e^{-\beta bq^2} dq \right)^N \end{aligned} \quad (4.47)$$

where in the last line I noticed that all the integrals over the p_i, q_i of each oscillator are mathematically identical.

The beauty of knowing the partition function is that we can now find the energy immediately, from the expression 4.41. It is

$$\langle E_s \rangle = -\frac{\partial (\ln Z)}{\partial \beta} . \quad (4.48)$$

For our expression above, we find

$$\begin{aligned}\langle E \rangle &= -N \frac{\partial}{\partial \beta} \left[\ln \left(\int e^{-\beta ap^2} dp \int e^{-\beta bq^2} dq \right) \right] \\ &= -N \frac{\partial}{\partial \beta} \left[\ln \left(\int e^{-\beta ap^2} dp \right) + \ln \left(\int e^{-\beta bq^2} dq \right) \right]\end{aligned}\quad (4.49)$$

which using $d(\ln x(y))/dy = (dx/dy)/x$ gives

$$\langle E \rangle = N \left[\frac{\int ap^2 e^{-\beta ap^2} dp}{\int e^{-\beta ap^2} dp} + \frac{\int bq^2 e^{-\beta bq^2} dq}{\int e^{-\beta bq^2} dq} \right]. \quad (4.50)$$

Now we note that the first term is the same as the second if a, p and b, q are exchanged. If we consider the first term only, we can make the substitution $\beta ap^2 = x^2$ and therefore $\sqrt{a}\sqrt{\beta}dp = dx$. Remembering that constants from the numerator and denominator can cancel, we find that the contribution from the kinetic energy term is then

$$\langle E(p) \rangle = \frac{N}{\beta} \frac{\int_{-\infty}^{\infty} x^2 e^{-x^2} dx}{\int_{-\infty}^{\infty} e^{-x^2} dx}. \quad (4.51)$$

Note that nothing here depends on a ! The integrals produce a numerical factor of $1/2$, so we finally get

$$\langle E(p) \rangle = \frac{N}{2\beta} = \frac{Nk_B T}{2}. \quad (4.52)$$

If we divide both sides by N , we find that for each particle, we get a contribution from the kinetic energy which is

$$\frac{k_B T}{2}. \quad (4.53)$$

Now note that this term does not depend on a . As a result, the contribution from the bq^2 will not depend on b (the integrals are identical), and therefore the potential energy terms of each particle will also contribute

$$\frac{k_B T}{2}. \quad (4.54)$$

That the contributions do not depend on the parameters of the energy dependence is striking. This is known as the equipartition theorem. In words, it states that for every degree of freedom which contributes a energy term which is quadratic, we get a contribution of $k_B T/2$ to the mean energy of a system in thermal equilibrium. It is worth emphasizing at this stage that the independence of the integral on the energy is peculiar to the particular form of the energy which we chose. For both of our phase-space variables the dependence of the energy on the variable was quadratic, and it is this which allows us to arrive at the equipartition theorem. The equipartition theorem is important because many energies have this form, for instance anything depending on kinetic energy, and any potential which can be expanded about equilibrium as a harmonic potential (which the Taylor expansion makes fairly likely).

A quick question for you: What would the mean energy of particles in a 3-D gas be if they were in equilibrium with a bath at temperature T ?

4.5.1 The equipartition theorem for molecules

If instead of a collection of point particles, we had a collection of molecules in thermal equilibrium, then there are extra terms in the energy which are also quadratic. This is because the internal structure of a molecule can also have energy.

For a two-atom molecule, we start by considering the translation of the center of mass. If the molecule is in free space, this will contribute $3k_B T/2$ due to kinetic energy. But the molecule can also rotate about the center of mass. For a two-atom molecule, it can do this in two directions. Both of these contribute a quadratic term to the energy, giving another $2k_B T/2$. Finally, the atoms can vibrate along the axis of the bond. The force between the atoms produces a potential similar to that of a harmonic oscillator, and the potential energy from the chemical bond plus the kinetic energy associated with the stretching motion of the bond give an additional $2k_B T/2$ to the mean energy. We now count these up and get 7 independent terms - the mean energy of a collection of N such molecules at temperature T is therefore predicted to be $7Nk_B T/2$.

People aren't generally able to measure the energy of a collection of atoms or molecules directly. A more common experiment is to do some work on the system, and observe the temperature rise. This allows the heat capacity to be measured, which is defined as

$$C = \frac{\partial \langle E \rangle}{\partial T} . \quad (4.55)$$

From the equipartition theorem, you would expect that this is constant (and indeed for a gas of atoms near room temperature it is). For the molecules discussed in the previous paragraph, we should find a value $C = 7Nk_B/2$. However what was observed in experiments performed on molecules was a dependence similar to that shown in figure 4.7. Here we see that in certain temperature ranges the heat-capacity is flat (at $3Nk_B/2$, $5Nk_B/2$, and $7Nk_B/2$), but there are also transition regions. Indeed if we think a bit harder about the equipartition theorem, we quickly realize that it must be wrong; a gas of molecules each consisting of a large number of atoms and thus a large number of terms contributing to the energy would have a huge heat capacity. This was realized by Maxwell very quickly after coming up with a basic notion of atoms as spherical objects and molecules as multiple spherical objects stuck together. A nice quote from Lenard in this regard is that certain types of energy seemed to be "frozen out" at low temperatures - this is indeed true, and the explanation for this was resolved during the formation of quantum mechanics.

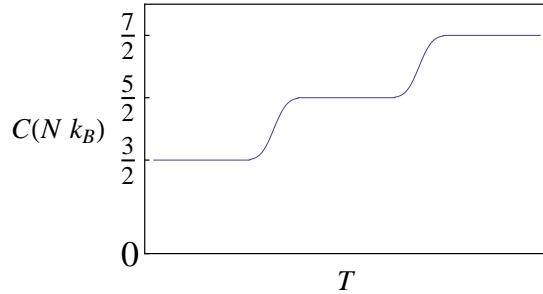


Figure 4.7: Schematic of the temperature dependence of the heat capacity for a diatomic molecule. The heat capacity is the derivative of energy with respect to temperature. For low temperatures the molecules act as though only the terms associated with motion of the center of mass were involved. For higher T , an extra Nk_B is added, which is due to "activation" of the rotational motion of the molecules. For even higher temperatures the molecule acts as if there is an additional Nk_B , which is due to "activation" of the vibrational motion. We shall describe what the "activation" involves later in the course.

4.5.2 Mean energy for a discrete system

Previously we considered a set of particles in a harmonic oscillator potential which had a continuous energy spectrum, and attempted to evaluate the mean energy of the oscillators.

We found that the mean energy per degree of freedom does not depend on the frequency of the oscillator. Let us return to this problem, but now we will only let the energy of each oscillator take discrete values $n\epsilon_i$, where n is an integer. In order that we will be able to consider oscillators of different frequencies, I have labeled the value of the energy of oscillator i by ϵ_i .

In order to calculate macroscopic variables for a collection of N such oscillators, we first look at the partition function for a single oscillator. According to our earlier working, this is given by the Boltzmann distribution.

$$Z_i = \sum_n^{\infty} e^{-\beta n \epsilon_i} = \frac{1}{1 - e^{-\beta \epsilon_i}} \quad (4.56)$$

If the oscillators are distinguishable, the partition function for many oscillators will be

$$Z = \prod_i^N Z_i \quad (4.57)$$

In order to find the mean energy of the collection of oscillators, we do the same as we did for the continuous oscillators, using $\langle E \rangle = -\partial(\ln Z)/\partial\beta$. We find

$$\begin{aligned} \langle E \rangle &= - \sum_i^N \frac{\partial(\ln Z_i)}{\partial\beta} \\ &= - \sum_i^N \frac{1}{Z_i} \frac{\partial Z_i}{\partial\beta} \\ &= \sum_i^N \frac{\epsilon_i}{e^{\beta \epsilon_i} - 1} \end{aligned} \quad (4.58)$$

Now we see very different behavior: the heat capacity is certainly not constant with respect to temperature. Note that the only thing we did differently here was to only allow the energy to take discrete values. We will see later that Planck was the first to take this step, which he found was necessary to explain the spectrum of radiation emitted from hot sources.

4.6 Blackbody experiments and concepts

In the 1800s, there were a range of experimental results which concerned the measurement of blackbody radiation. These presented a fundamental challenge to physicists, who had grave problems making theories which fit the data. This was eventually done by Planck.

All objects absorb and emit radiation, and depending on their environment the spectrum of the light emitted will vary. The special feature of blackbody radiation is that the object which emits radiation is in thermal equilibrium at a uniform temperature T , and that it absorbs all radiation which is incident on it. One possible blackbody apparatus consists of a closed oven where the walls are held at temperature T and there are no windows for light to get out (see figure 4.8). The oven walls will absorb and emit radiation continuously, and in the long-time limit we would expect the walls of the oven to come into thermal equilibrium with the light which is bouncing around inside the oven. We could observe the spectrum of light in the oven by poking a small hole in the side and letting a small amount of the light escape. The spectrum could be measured using some kind of spectrometer.

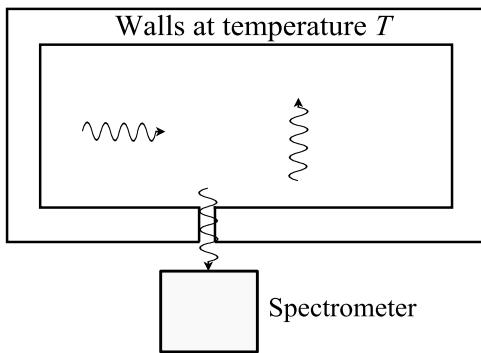


Figure 4.8: Experimental setup imagined in the consideration of black-body radiation. The radiation inside the box is in thermal equilibrium with the walls of the cavity, which are maintained at temperature T . The spectrum of radiation inside the box is measured by making a small hole in the box, and directing light emitted through the hole into a spectrometer.

The important feature of blackbody radiation is that so long as an object which absorbs all light incident on it is in thermal equilibrium at temperature T , the spectrum of radiation which it emits will be the same. So shouldn't we see light coming from a lump of charcoal at room temperature? There *is* electromagnetic radiation emitted, but it is not in the visible region of the spectrum. We shall see later that the spectrum of this light is in the infra-red region of the spectrum.

The oven above had absorbing walls which brought the light inside into thermal equilibrium. What if we have a box with fully reflective sides, where we know that the radiation inside is also in thermal equilibrium at temperature T . Will it have the same spectrum? The answer is yes! If not, we could place the two boxes (absorbing and reflective) next to each other, and open a hole up between the two. If the spectra are different, then the energy of the radiation fields will be different, and it will be possible to do work without having a difference in temperatures by moving radiation between boxes. This would violate the second law of thermodynamics.

The reflective box is much easier to treat than the absorptive oven, so this is the case we will describe here. We will however make one step, which is to assume that the light fields in the box behave like a harmonic oscillator. It turns out that this is an excellent assumption - in quantum mechanics, the key steps in quantizing the behavior of light is to notice that

the energy of light looks very similar to that of a harmonic oscillator.

4.6.1 Optical modes of a cavity

Let us consider the radiation inside a closed box of dimensions L_x, L_y, L_z coated on the inside with reflective surfaces. If we remember our boundary conditions for electric and magnetic fields at conducting surfaces, we know that the electric field parallel to the surface should be zero, and the magnetic field perpendicular to the surface should also be zero.

Solutions to the wave equation with these boundary conditions are standing waves which have normal mode solutions of the form

$$E(x, t) = A \sin(k_x x) \sin(k_y y) \sin(k_z z) \sin(ckt) \quad (4.59)$$

where $k_x = n_x \pi / L_x$ and $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$. n_x, n_y and n_z are positive integers which tell us how many standing wave periods fit within the box for each of the directions. For every value of k , we can choose multiple combinations of n_x, n_y and n_z , each of which correspond to different configurations of fields in the cavity. If these field configurations do not interact with each other, then each of them can be treated as an independent oscillation mode of the system. We then have to work out how many types of mode we have at each value of k .

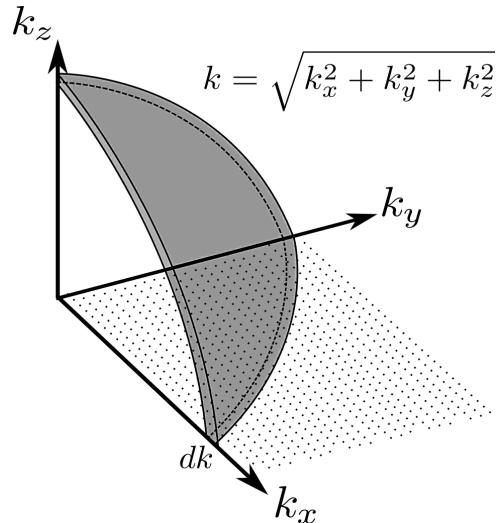


Figure 4.9: In a space defined by (k_x, k_y, k_z) each mode sits at a point which is some integer multiple of the spacing of modes in a given direction. These spacings are $(\pi/L_x, \pi/L_y, \pi/L_z)$ in the respective directions. The volume of each mode is $\pi^3/(L_x L_y L_z)$. The number of modes between k and $k + dk$ is given by those for which the points fall within the shaded volume.

To count the modes, we imagine a discrete 3-D space made up of a grid of points at every value of (k_x, k_y, k_z) for which a mode exists (see figure 4.9). This is called k -space. The distance between neighboring mode points in the k_x direction is π/L_x , while for k_y it is π/L_y , and similarly for k_z . Therefore we can assign a volume $\pi^3/(L_x L_y L_z)$ for each mode point. In this space, the volume of the modes between k and $k + dk$ is given by a spherical shell in the positive quadrant (this is shown in figure 4.9) which has volume

$$\frac{1}{8} 4\pi k^2 dk . \quad (4.60)$$

The number of modes is given by this volume divided by the volume of each mode, namely

$$\frac{\frac{1}{8} 4\pi k^2 dk}{\pi^3/(L_x L_y L_z)} = \frac{V 4k^2 dk}{8\pi^2} \quad (4.61)$$

where $V \equiv L_x L_y L_z$ is the volume of the box in real space. Since $k = 2\pi\nu/c$, we can rewrite this as a number of modes with frequency between ν and $\nu + d\nu$ which gives

$$g(\nu)d\nu = \frac{4\pi\nu^2 V}{c^3}d\nu . \quad (4.62)$$

Actually this is not quite correct. We know from earlier in our course that for a single spatial mode, the light can have two orthogonal polarizations where the electric field is parallel to the walls (horizontal and vertical). We must therefore multiply the result we found above by a factor of 2 to account for this. The final result for the number of modes with frequency between ν and $\nu + d\nu$ is therefore

$$g(\nu)d\nu = \frac{8\pi\nu^2 V}{c^3}d\nu . \quad (4.63)$$

4.6.2 Classical and quantum distributions

Rayleigh-Jeans law

Now we are ready to find the energy density in the box. Let us assume that each mode is independent (ie. that they do not interact with each other) but nevertheless that they are in thermal equilibrium at temperature T . We will treat each mode as if it were a harmonic oscillator.

Each of the oscillators has a well defined frequency. If each mode also has a continuous distribution of energies (which if the frequency is fixed is dependent on the amplitude of oscillation), then each of the harmonic oscillators will have a mean energy $\langle E \rangle = k_B T$. From the previous section, we know the number of oscillators with frequency between ν and $\nu + d\nu$, so we find that the energy density (energy per unit volume) in the spectral region between ν and $\nu + d\nu$ is given by the *Rayleigh-Jeans law* (from a series of derivations performed between 1900-1905)

$$\rho(\nu)d\nu = k_B T \frac{g(\nu)d\nu}{V} = k_B T \frac{8\pi\nu^2}{c^3}d\nu \quad (4.64)$$

For low frequencies, this result reproduces the behavior which was observed in the late 1800s in experiments. However the result is clearly wrong, because it implies that at high frequencies, there will be higher and higher energy densities. This was called the *ultraviolet catastrophe* and was recognized as a severe problem for classical physics.

Planck distribution

Planck instead allowed only energy values which are integer multiples of $h\nu$, where h is called Planck's constant and ν is the frequency of the oscillator. From equation 4.58 earlier in the lecture course we have

$$\langle E \rangle = \sum_i^N \frac{\epsilon_i}{e^{\beta\epsilon_i} - 1} . \quad (4.65)$$

which means that the energy contribution from a single oscillator with frequency ν is

$$\langle E_{SO}(\nu) \rangle = \frac{h\nu}{e^{\beta h\nu} - 1} \quad (4.66)$$

The sum over all the oscillators will count a large number of oscillators with frequency between ν and $\nu + d\nu$, where this number is given by equation 4.63. By putting these two

contributions together, we arrive at the famous Planck distribution

$$\rho(\nu)d\nu = \frac{\langle E_{SO}(\nu) \rangle g(\nu)d\nu}{V} = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{\beta h\nu} - 1} d\nu = \frac{8\pi h}{c^3} \frac{\nu^3 d\nu}{e^{\beta h\nu} - 1} \quad (4.67)$$

This is the spectrum which was observed in the experiments, and is one of the most striking results pointing to the fact that the energy of an oscillator is only available to it in discrete lumps, which became known as quanta.

The high-temperature limit of the Planck distribution is worth examining. For $\beta = 1/(k_B T) \ll 1/(h\nu)$ (or rather $k_B T \gg h\nu$), we find

$$\frac{1}{e^{\beta h\nu} - 1} \simeq \frac{1}{1 + \beta h\nu - 1} = \frac{k_B T}{h\nu} \quad (4.68)$$

which allows us to recover the Rayleigh-Jeans result

$$\rho(\nu)d\nu = k_B T \frac{8\pi\nu^2}{c^3} d\nu \quad (4.69)$$

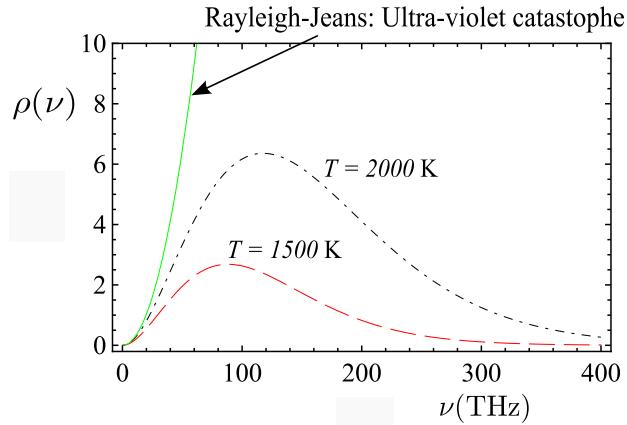


Figure 4.10: Energy spectral density per unit volume given by the Planck distribution of blackbody radiation, at temperatures $T = 1500$ K (red, dashed) and $T = 2000$ K (black, dot-dashed). The units are $\text{J}/\text{Hz}/\text{m}^{-3}$. Also shown is the Rayleigh-Jeans distribution for $T = 2000$ K (solid, green), illustrating the “ultraviolet catastrophe” which was such an unnerving prediction for classical physics.

Wien displacement law

One interesting result which can be seen from the Planck distribution is that the frequency with the highest energy density is proportional to the temperature. We can see this by looking for the maximum using differentiation

$$\frac{d\rho(\nu)}{d\nu} = 0 \quad (4.70)$$

which we can show gives a condition on $x = h\nu_{\text{maximum}}/(k_B T)$ of

$$3 = (3 - x)e^x . \quad (4.71)$$

This has a solution (numerically found) of $x = 2.822$. The dependence of the maximum on temperature was actually noticed by Wien prior to Planck finding his distribution, from careful considerations which we shall not go into here.

Stefan-Boltzmann law

It is often important to know how much radiation is emitted from a surface, when integrated over all frequencies. This is clearly impossible in the Rayleigh-Jeans picture, because the value of this integral will tend to infinity. For the Planck distribution, we can however take the integral over ν to find the total energy density

$$u(T) = \int_0^\infty \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/(k_B T)} - 1} d\nu \quad (4.72)$$

from which the temperature dependence can be deduced by making the substitution $x = h\nu/k_B T$, giving

$$u(T) = \frac{8\pi k_B^4 T^4}{h^3 c^3} \int_0^\infty \frac{x^3}{e^x - 1} dx \quad (4.73)$$

$$= aT^4 \quad (4.74)$$

This is called the *Stefan-Boltzmann* law. The integral over x is not the simplest to perform, evaluating to give $\int_0^\infty x^3/(e^x - 1)dx = \pi^4/15$. The constant a is most often quoted in terms of another constant $\sigma = 5.670 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ as $\sigma = ca/4$.

4.6.3 Einstein and the Planck Distribution

A more rigorous derivation of the Planck distribution was provided at a later stage by Einstein. Around the same time Einstein provided an explanation of the photo-electric effect in terms of “quanta” – discrete lumps of light. He combined these quanta with the notion of discrete energy levels in atoms, and came out with the spectrum of blackbody radiation - this was a quite fantastic piece of work which paved the way to quantum mechanics.

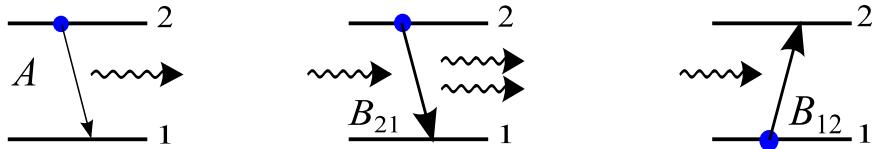


Figure 4.11: Three possible processes which can cause a transition between energy levels 1 and 2 of an atom. Left: the atom spontaneously decays from level 2 to level 1, emitting a photon into a random direction, but with a frequency corresponding to the energy difference between the levels. This happens at a rate A for a single atom. Center: the atom is stimulated to make a transition from 2 to 1 by interaction with a photon already in the field, and emits a photon of exactly the same type which caused it to decay. The rate of transitions of this type is $B_{21}\rho(\nu)$ Right: an atom starting in level 1 is stimulated to absorb a photon. This happens with a rate given by $B_{12}\rho(\nu)$.

The Einstein derivation deals with cavity modes as before, but which interact with atoms in the walls of the cavity. The atoms are able to absorb and emit light only in discrete lumps, and only one at a time. When the atom absorbs a quantum of light, it makes a transition between two discrete energy levels, which we will call level 1 and level 2. Einstein postulated that light is emitted and absorbed in three ways.

- An atom starting in level 2 can decay. In doing so, it will emit one quanta of light with frequency $\nu = (E_2 - E_1)/h$. For N_2 atoms in level 2, this spontaneous emission process will occur at a rate AN_2 .
- An atom starting in level 2 interacts with a photon already in the light field, causing the atom to make a transition to level 1, and to emit an additional photon at frequency

ν . The rate of such stimulated emission for N_2 atoms starting in level 2 is given by $B_{21}\rho(\nu)$

- An atom in level 1 can absorb a photon from the light field. The rate of absorbtion for N_1 atoms in level 1 is $B_{12}\rho(\nu)$

We can now write down rate equations for the populations of the levels. For N_2 , this looks like

$$\frac{dN_2}{dt} = -AN_2 + \rho(\nu)(B_{12}N_1 - B_{21}N_2) . \quad (4.75)$$

Once the system has arrived in thermal equilibrium, we know that $dN_2/dt = 0$ (this is a static steady state). Thus we find

$$\rho(\nu) = \frac{A}{B_{12}N_1/N_2 - B_{21}} \quad (4.76)$$

In thermal equilibrium, we also know from Boltzmann's law that the relative probability that two levels with energy separation $h\nu$ are occupied is given by $e^{-h\nu/(k_B T)}$. Therefore

$$\rho(\nu) = \frac{A}{B_{12}e^{h\nu/(k_B T)} - B_{21}} \quad (4.77)$$

By comparison with the Planck distribution, this formula allows us to find relations between the rate coefficients in the atom

$$B_{12} = B_{21} \quad (4.78)$$

$$A = \frac{8\pi h\nu^3}{c^3} B_{12} \quad (4.79)$$

This was a quite remarkable piece of work. Starting from some postulates about the nature of absorbtion and emission, we can verify the form of the Planck distribution with regard to temperature, and find the relative rates of different processes which go on in the interaction between light and matter. The Einstein coefficients A and B are extremely important in the internal dynamics of lasers, which since these are such important devices you will surely see again.

4.7 Indistinguishable particles

Up until now, we always considered distinguishable particles, where usually we distinguished them by location. Though valid for atoms in a solid, or independent modes of a blackbody cavity, this is not valid for identical atoms in a gas. If we start with a gas of identical atoms in a particular microstate, and swap two of the atoms, we end up with the same microstate.

To illustrate why this causes a problem, let us consider the case of putting two objects A and B into two boxes L and R . If A and B are distinguishable, the microstates of the system are $A_L B_L$, $A_R B_R$, $A_L B_R$ and $A_R B_L$. There are therefore 4 microstates in total. For the case where A and B are indistinguishable, the microstates $A_L B_R$ and $A_R B_L$ are identical - even in the microscopic description there is no way of telling the difference between the two. This means that we end up over-counting microstates where the two particles are in different states by a factor two.

4.7.1 Review: relation between partition functions for distinguishable particles

At this point it is also worth considering the relation between the single particle partition function Z_{SP} of equation 4.26 which we derived for distinguishable particles, and the partition function of the system Z given by equation 4.37 (repeated here using the definition $\beta = 1/(k_B T)$)

$$Z = \sum_s e^{-\beta E_s} \quad (4.80)$$

which is summed over all the microstates of the system, which have energy given by E_s . The latter is a weighted sum over all the microstates. A microstate is denoted by the state of each particle. E_s can therefore be written as a sum over the energies of these individual states, which is written for the state s as

$$E_s = \sum_k e_s^{(k)} \quad (4.81)$$

where $e_s^{(k)}$ is the energy of particle k when it is part of the microstate s . The equation for the partition function can then be written as

$$Z = \sum_s e^{-\beta e_s^{(1)}} e^{-\beta e_s^{(2)}} \dots e^{-\beta e_s^{(N)}} . \quad (4.82)$$

Now let us consider that the sum is over *all* microstates of the system. The total will therefore sum over products of all particles in all of their different states. If we label the i th state of the k th particle by i_k , the summation can be written as a sum over the individual states that the N individual particles can have and the partition function becomes

$$Z = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} e^{-\beta e_{i_1}^{(1)}} e^{-\beta e_{i_2}^{(2)}} \dots e^{-\beta e_{i_3}^{(N)}} = \sum_{i_1} e^{-\beta e_{i_1}^{(1)}} \sum_{i_2} e^{-\beta e_{i_2}^{(2)}} \dots \sum_{i_N} e^{-\beta e_{i_N}^{(N)}} \quad (4.83)$$

which if the particles are identical (ie. each sit in a potential which looks the same) but are *distinguishable* gives us

$$Z = Z_{SP}^N . \quad (4.84)$$

We shall see that the primary difference between the behavior of distinguishable and indistinguishable particles is that for indistinguishable particles it is *not* possible to relate the partition function of the system to that of the individual particles according to this expression. This affects any quantity which measures the number of states. Since the entropy is a measure of the number of microstates, this is the thermodynamic quantity which is primarily affected.

4.7.2 Counting states of indistinguishable particles

Since the methods we used previously don't work for indistinguishable particles, we must change our approach. A rigorous treatment of this problem would use something called the Gibbs ensemble. This is not what we will do here, because it would take more than one lecture, and one lecture is all we can afford if we are to complete our whole course by the end of the semester. Instead what I will offer is a derivation which recovers the same results as the rigorous treatment, but where the derivation is only strictly valid in a particular parameter regime. In this section, we will use this approach to derive the partition function of a classical gas. This will enable us to obtain the famous Maxwell-Boltzmann distribution of velocities in such a gas. Towards the end of the course we shall return to this approach and use it to derive properties of quantum gases. The important difference is not only quantized energy levels, but actually that *quantum particles can be indistinguishable in different ways*. This is one of the many counter-intuitive properties of quantum mechanics, which we find counter-intuitive because we all grew up in a "classical" world.

The situation we will consider here is a set of states with very closely-spaced energy levels. A particle can be placed into any one of these states. We will make the approximation that a number g_i of states which are very close to each other in energy can be put together into a bundle with a single energy ϵ_i . Now we count the number of ways in which n_i particles can be assigned to the bundle, which will give the number of microstates of the bundle in which the g_i states contain n_i particles. We will use this number to find the number of microstates of the system.

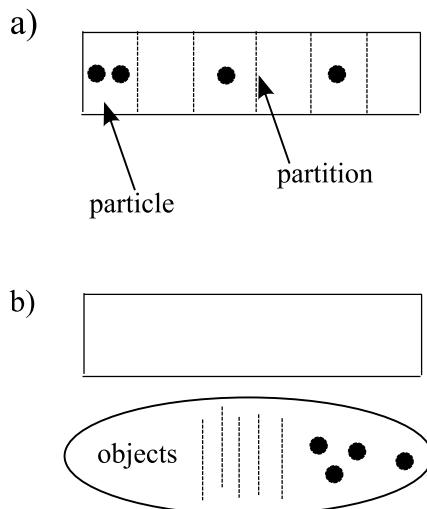


Figure 4.12: a) an example of a microstate of a bundle with $g_i = 6$ states and $n_i = 4$ particles. There are $g_i - 1$ partitions between the states. b) We can find the number of microstates of this bundle with a given g_i, n_i by treating both the partitions and the particles as "objects" and asking about the number of orders in which we can place the objects into the box. For the example of part a), working from left to right, we would need to pick particle, particle, partition, partition, particle, partition, particle, partition to get the microstate.

Consider that the g_i states are boxes. There will be $g_i - 1$ partitions between them (an example with $g_i = 6$ is given in figure 4.12 a)). We will imagine these partitions as "objects", and include the particles as "objects" too. Then let's start in the first box. We can move to the second box by picking any one of the partitions, or we can choose to put a particle in this box by picking one of the particles. There are a total of $n_i + g_i - 1$ ways to choose an object. After we have chosen the first one, there are $n_i + g_i - 2$ ways to choose the second object, and so on. The total number of possibilities for choosing the whole lot is $(n_i + g_i - 1)!$.

However, for the particles, we cannot tell between the $n_i!$ different orders in which we might have chosen them. The same is true for the $(g_i - 1)!$ orders in which I could have chosen the partitions. Therefore the total number of configurations for this bundle (where all states have energy ϵ_i) is

$$\Omega_i = \frac{(n_i + g_i - 1)!}{n_i!(g_i - 1)!} . \quad (4.85)$$

The total number of microstates for the system is the product of those for the different ϵ_i , and is thus

$$\Omega = \prod_i \frac{(n_i + g_i - 1)!}{n_i!(g_i - 1)!} . \quad (4.86)$$

4.7.3 The classical limit - single particle partition function

Now we will consider the *classical* limit, where the number of states g_i is assumed to be very large compared to n_i for all i . The consequence of this is that it is very unlikely that two particles will be in the same state. For one mole of a typical gas at room temperature, the total number of states is around 10^{30} and the total number of atoms is 6×10^{23} . If we split the total energy scale of the atoms into 10^{10} parts then we will have 10^{10} bundles. This means that $g_i \sim 10^{20}$ and $n_i \sim 10^{13}$ the approximation is good. The other approximation we made earlier is that each state in a bundle has the same energy. Since we split the energy scale into 10^{10} parts, we have good reason to believe that this is also a good approximation.

In the limit $g_i \gg n_i$, we can use Stirling's theorem to simplify the expression for Ω

$$\ln \left(\frac{(n_i + g_i - 1)!}{n_i!(g_i - 1)!} \right) \simeq -\ln(n_i!) + n_i \ln(g_i) \quad (4.87)$$

and therefore

$$\Omega \simeq \prod_i \frac{g_i^{n_i}}{n_i!} \quad (4.88)$$

Now that we have the number of states, we can proceed to look for the equilibrium macrostate by maximizing the number of states over the n_i . As before, we work under the constraint that the gas has a fixed total energy and total number of particles. Thus we need to satisfy $\sum_i n_i \epsilon_i = E$ and $\sum_i n_i = N$ and introduce Lagrange multipliers α and β such that the maximization is

$$0 = \frac{\partial}{\partial n_i} \left(\ln(\Omega) - \alpha \left(\sum_i n_i - N \right) - \beta \left(\sum_i n_i \epsilon_i - E \right) \right) \quad (4.89)$$

which using Stirling's theorem again gives

$$0 = \ln g_i - \ln n_i - \alpha - \beta \epsilon_i \quad (4.90)$$

and

$$n_i = g_i e^{-\alpha - \beta \epsilon_i} = A g_i e^{-\beta \epsilon_i} \quad (4.91)$$

where A is obtained from

$$N = A \sum_i g_i e^{-\beta \epsilon_i} \quad (4.92)$$

The probability for a selected particle to be in a state with energy ϵ_i we recover a Boltzmann distribution again

$$p_i = \frac{g_i e^{-\beta \epsilon_i}}{\sum_i g_i e^{-\beta \epsilon_i}} = \frac{g_i e^{-\beta \epsilon_i}}{Z_{SP}} \quad (4.93)$$

which is the same result we found above (if you're worried about the g_i , you should not be, because formerly I was looking at the probability to be in the state i , and now I am looking at the probability that the particle is in a state with energy ϵ_i).

4.7.4 The many-particle partition function

You might be thinking, why did I bother? Since all states with a given i above have the same energy ϵ_i , I have merely recovered equation 4.40 with a different set of constants (sorry about the notation change). The reason that it looks like this is that I made a jump back to looking at a single particle again. Indistinguishability affects the number of microstates of the system, not the occupancies of the single particle states (at least at the level where no two particles occupy the same state). This is a crucial difference. The major difference is the entropy, which measures the number of possible microstates.

The entropy of the gas of indistinguishable particles, as for the distinguishable case, is defined by

$$S = k_B \ln(\Omega) \quad (4.94)$$

where Ω is the number of microstates of the system. For the indistinguishable case, we found

$$\Omega_I = \prod_i \frac{g_i^{n_i}}{n_i!} \quad (4.95)$$

and therefore

$$\ln(\Omega_I) = \sum_i (n_i \ln(g_i) - \ln(n_i!)) \quad (4.96)$$

$$\simeq \sum_i (n_i \ln(g_i) - n_i \ln(n_i) + n_i) \quad (4.97)$$

$$\simeq \sum_i n_i \ln(g_i/n_i) + N \quad (4.98)$$

where for the second line I again used Stirling's rule. We now make use of equation 4.91, re-arranging to find

$$\frac{g_i}{n_i} = \frac{1}{A} e^{\beta \epsilon_i} , \quad (4.99)$$

which allows us to write equation 4.98 as

$$\ln(\Omega_I) \simeq \sum_i n_i (-\ln(A) + \beta \epsilon_i) + N \quad (4.100)$$

We now make use of the fact that

$$\sum_i n_i = N \quad (4.101)$$

which, from equation 4.91 gives us that

$$N = A Z_{SP} \quad (4.102)$$

with Z_{SP} the single particle partition function. As a result we can use $\ln(A) = \ln(N) - \ln(Z_{SP})$ to write

$$S_I = k_B \ln(\Omega_I) = -k_B N \ln(N) + k_B N \ln(Z_{SP}) + \frac{\langle E \rangle}{T} + k_B N \quad (4.103)$$

$$= k_B N \ln(Z_{SP}) - k_B \ln(N!) + \frac{\langle E \rangle}{T} \quad (4.104)$$

This differs from the value for the distinguishable case by the factor $k_B \ln(N!)$. Therefore where in the case of distinguishable particles the partition function is

$$Z_D = Z_{SP}^N \quad (4.105)$$

for the case of indistinguishable classical particles we find that

$$Z_I = \frac{Z_{SP}^N}{N!} . \quad (4.106)$$

4.7.5 Back to the ideal classical gas.

The ideal classical gas follows exactly the working above. For a single particle, we have that the kinetic energy is given by $\epsilon = p^2/(2m)$. We therefore only need to know the degeneracy g_i in order to find the single-particle partition function. We use a method similar to those used for the modes of the blackbody cavity above. Now we are talking about states though, rather than modes as we were before. We want to find the number of states with momentum between p and $p+dp$. Since states of the particle are specified by both positions and momenta we can find the volume of phase space associated with such states

$$\int_V dx dy dz \int_{p, p+dp} d\mathbf{p} = V \int_{p, p+dp} d\mathbf{p} \quad (4.107)$$

where V is the volume of the gas and \mathbf{p} is a vector in a three dimensional space defined by the momenta. As with the case where we counted modes in k -space, the integral over momentum is given by the volume of a shell of thickness p , only now since p_x, p_y, p_z can be positive and negative we take the volume over the whole spherical surface, not just the positive quadrant. This volume is

$$4\pi p^2 dp \quad (4.108)$$

in momentum space, and the volume in phase space is

$$V 4\pi p^2 dp . \quad (4.109)$$

To find the number of states, we now need to define the volume of a single state in this same space. This is arbitrary. Note that the units are

$$(\text{kgms}^{-1})^3 \text{m}^3 \quad (4.110)$$

so we need to cube a quantity with dimensions of mass times length squared divided by time. This is the same type of units as Planck's constant, J s. You will very often see Planck's constant used (in fact this is the way in which Planck's constant was first introduced), however, since this is a classical gas, you'd be right to be a bit perplexed. I will show below that there are no physical consequences to this choice, so don't worry about this. If we do use Planck's constant, we find that the number of states with momentum magnitude between p and $p+dp$ is

$$f(p)dp = \frac{4\pi p^2 V dp}{h^3} . \quad (4.111)$$

We are now ready to find the partition function for a single particle in the ideal gas. This is given by

$$Z_1 = \int_0^\infty f(p) e^{-p^2/(2m)} dp = \int_0^\infty \frac{4\pi p^2 V dp}{h^3} e^{-\beta p^2/(2m)} \quad (4.112)$$

$$= V \left(\frac{2\pi m k_B T}{h^2} \right)^{3/2} \quad (4.113)$$

and thus that the partition function for the whole gas if it contains N particles is

$$Z = \frac{Z_1^N}{N!} = \frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h^2} \right)^{3N/2} \quad (4.114)$$

Note that this contains none of the microscopic properties of the gas, apart from the mass of the particles. Otherwise, we've managed to write things in terms of macroscopic variables, such as temperature, volume and particle number.

Let us return to our point about h . As an example of the independence of physical quantities on h , let us consider the mean energy. As we learnt earlier in the course, the mean energy is

$$U = -\frac{\partial(\ln Z)}{\partial\beta} = -\frac{\partial}{\partial\beta} \left[N \ln V - \ln N! - \frac{3N}{2} \ln \beta + \frac{3N}{2} \ln(2\pi m/h^2) \right] \quad (4.115)$$

$$= \frac{3Nk_B T}{2}. \quad (4.116)$$

If we think harder, we should realize that a constant in the density of states should not affect quantities such as energy – these come from equations for probabilities for finding states of given energy occupied, and these typically look like

$$p_E dE = \frac{g(E) dE e^{-\beta E}}{\int g(E) dE e^{-\beta E}}. \quad (4.117)$$

Clearly, any constant in the density of states will factor out.

4.7.6 The Maxwell-Boltzmann velocity distribution

We will close this part of the course by looking at one of the most useful distributions of classical statistical mechanics, namely the velocity distribution of an ideal gas of particles. This is famous because it is one of the first distributions which was rigorously measured, using the apparatus shown in figure 4.13. Here atoms are allowed to escape from an oven held at temperature T , and pass through two discs rotating at angular frequency ω . These discs have slits cut into them. Particles can only pass through both slits, if they have a velocity which satisfies $D/v = \delta\phi/\omega$ where $\delta\phi$ is the phase delay between the two discs. By varying the phase delay between the two discs, the relative number of atoms with different velocities can be measured.

The Maxwell-Boltzmann distribution gives the probability of finding an atom inside the box with a speed u (the average speed of those measured by the experiment described above is subtly different, as you will find out in the problem set). From the partition function for the ideal gas combined with the Boltzmann distribution, we can see that the probability of finding an atom with momentum magnitude between p and $p+dp$ is

$$P(p) dp = f(p) dp \frac{e^{-\beta p^2/(2m)}}{Z_1} \quad (4.118)$$

$$= \frac{4\pi p^2 dp}{h^3} \left(\frac{h^2}{2\pi m k_B T} \right)^{3/2} e^{-p^2/(2m k_B T)} \quad (4.119)$$

$$= 4\pi p^2 dp \frac{1}{(2\pi m k_B T)^{3/2}} e^{-p^2/(2m k_B T)} \quad (4.120)$$

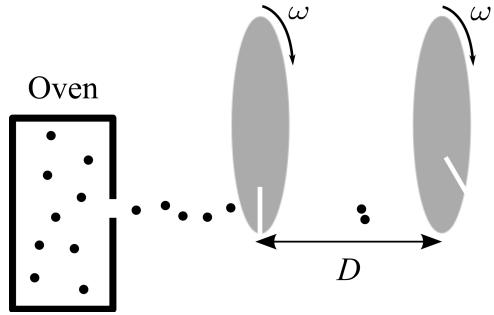


Figure 4.13: Apparatus to measure Maxwell's velocity distribution. Atoms effuse from an oven in a beam towards disks rotating with angular frequency ω . Each disk has a slot cut in it. At every time where the slot in the first disk is aligned with the beam, atoms get through. These atoms travel towards the second disk, and are passed if the time of arrival at this disk is the same as the time at which the slot points vertically down. The number of atoms which pass both disks is recorded by a detector on the right hand side (not shown).

and therefore the number of particles in the gas with velocities between u and $u + du$ is

$$n(u)du = N4\pi u^2 du \left(\frac{m}{2\pi k_B T} \right)^{3/2} e^{-mu^2/(2k_B T)} \quad (4.121)$$

Examples of the Maxwell-distribution are shown in figure 4.14 for three different temperatures. As would be expected, the distribution moves towards higher velocities for higher temperatures. The mean velocity can of course be easily found, using

$$\frac{\int_0^\infty un(u)du}{N} , \quad (4.122)$$

which for the Maxwell distribution gives

$$\langle u \rangle = \sqrt{\frac{8k_B T}{\pi m}} . \quad (4.123)$$

This is noticeably different from the position of the maximum in the distribution. This occurs when

$$\frac{dn(u)}{du} = 0 , \quad (4.124)$$

which leads to the expression

$$2u - u^2 \frac{mu}{k_B T} = 0 \quad (4.125)$$

and thus

$$u_{\max} = \sqrt{\frac{2k_B T}{m}} . \quad (4.126)$$

The ratio of these two is therefore $\langle u \rangle / u_{\max} = 2/\sqrt{\pi} = 1.13$.

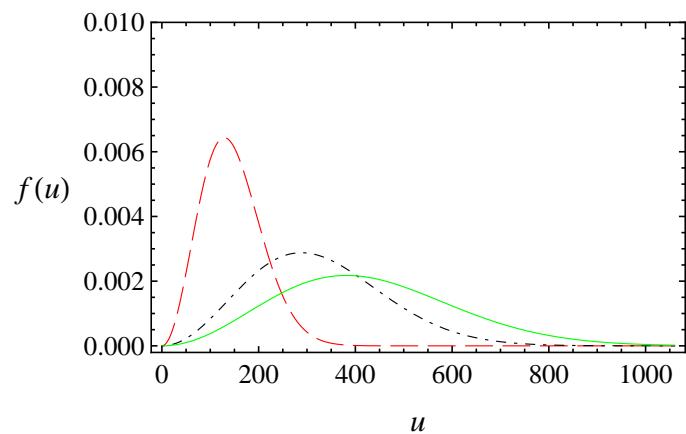


Figure 4.14: Maxwell distribution of velocities for a calcium gas at three different temperatures. These are 40 K: red, dashed, 200 K: black, dot-dashed and 300 K: green, solid curve. The units for u are ms^{-1} .

Chapter 5

Wave-particle “duality”

At this point in the course we shall start to consider that what we usually consider as particles can have wave properties, and vice versa. We have already seen that for light different experiments turn up different types of behavior. In optics, we saw that light *diffraction* if we try to pass it through a small aperture, and that light which is split up and travels along different paths *interferes* when it is recombined. These are the classic signatures of waves. On the other hand we all heard the discrete clicks from the photo-multiplier tube, and we learnt about Compton scattering. These are signs of light acting more like a tennis ball, ie. that it behaves in the same way as a particle.

5.1 Matter waves

5.1.1 The Davisson-Germer experiment

The first clear experimental evidence for the wave nature of particles other than light was produced by Davisson and Germer. Davisson and Germer slowed electrons from a tungsten filament source down to low velocities, and observed the number of electrons scattered by a Nickel crystal at a range of angles. They saw a striking feature of wave behavior, namely interference. That is, they observed maxima and minima in the number electrons reflected as a function of angle. The width of the maxima in the interference decreased as the energy of the electrons was increased.

5.1.2 de Broglie hypothesis

To explain these experiments, de Broglie hypothesized that all particles can have wave-like character. He assigned a characteristic wavelength to a particle which is called the de Broglie wavelength, and is given by

$$\lambda = \frac{h}{p} \quad (5.1)$$

where h is Planck's constant and p is the momentum of the particle. This form was inspired by the relation between momentum and wavelength for photons. There we found that for light of frequency f , the momentum is given by

$$p = \frac{hf}{c} = \frac{h}{\lambda} \quad (5.2)$$

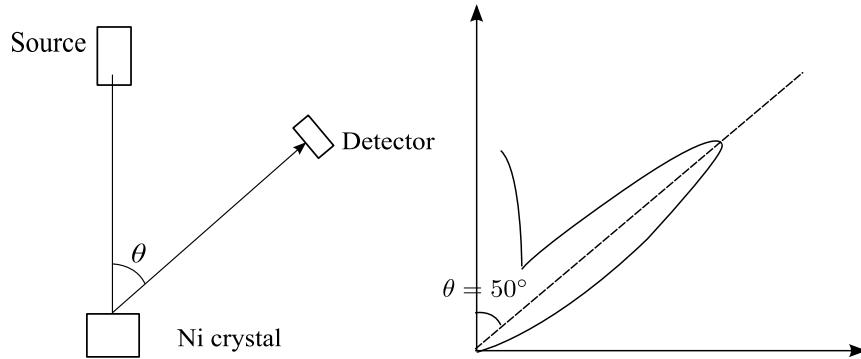


Figure 5.1: Left: Experimental setup in the experiment of Davisson and Germer. Electrons from a low-energy source are directed on a Nickel crystal target. Electrons scattered from the target are detected at an angle θ . Right: experimental results. The distance of the curve from the origin at a given angle is proportional to the intensity of the electrons scattered at that angle. A noticeable peak is observed at an angle of 50° . This is due to a form of Bragg scattering, though in the case of electrons the scattering is from the surface of the material (low-energy electrons do not penetrate far into solids).

So why is it not common to see diffraction effects with particles? Here we should remember the form of our diffraction pattern from a slit of width Δ_x . We found that the distribution of intensity as a function of angle is

$$I(\theta) = I(0) \frac{\sin^2(\pi \sin \theta \Delta_x / \lambda)}{(\pi \sin \theta \Delta_x / \lambda)^2} \quad (5.3)$$

The first zero in this function as a function of angle occurs for $\sin \theta = \lambda / \Delta_x$. Here we see immediately see that for significant intensity to be found at a range of angles (diffraction) we require $\Delta_x \sim \lambda$. As a result, we should examine the wavelength of particles such as electrons, neutrons etc. to find out what length-scales are required for diffraction effects. Typical energies for these particles are a few to many electron-volts. Using classical mechanics we can relate this to the momentum of the particle

$$p = \sqrt{2mE} . \quad (5.4)$$

For an electron of energy 1 eV the momentum is therefore 5.4×10^{-25} kg m s $^{-1}$, giving a de Broglie wavelength of 1.2 nm. This is smaller than the wavelength of visible light by more than a factor of 100. We clearly cannot make slits which are this small. Thus it is the small size of the wavelengths of matter, a scale which is defined by the momentum and Planck's constant, which result in us not seeing diffraction in our everyday lives.

The reason that Davisson and Germer could observe interference with the electrons in their experiments is that the atoms in the Nickel crystal are spaced by distances of 0.215 nm, which is close to the de Broglie wavelength of 0.167 nm for the 54 eV electrons which they used. The interference that they observed is due to a form of Bragg-scattering. This is slightly different to the Bragg-scattering that we saw with X-rays, because electrons only scatter from the few layers of atoms at the surface of the crystal. This is due to the fact that the electron scattering cross-section is large for these relatively low-energy electrons - they do not penetrate far into matter.

5.1.3 Modern applications.

Remember from our study of optics that the resolving power of an optical instrument such as a telescope or microscope was limited by diffraction effects. For a microscope, the condition

was that the minimum distance between two objects should be $\lambda/(NA)$, where NA is the numerical aperture. It is clear that by reducing the wavelength, smaller objects can be resolved. In this regard, electrons and neutrons, with wavelengths on the nanometer scale, play a huge role in modern imaging throughout the physical and biological sciences. The most famous example is the electron microscope, but investigations of crystal structure by neutron scattering are also worth discussing.

The electron microscope

The electron microscope involves electrons which are either reflected or transmitted through the specimen. In its original form, the electrons were first emitted from a Tungsten filament cathode, and then accelerated by an anode up to energies of around 100 keV. These electrons are then focussed by electrostatic and electromagnetic optics: electrodes and magnetic loops which create electric and magnetic fields capable of focussing a beam of charged particles. The electron beam then passes through the specimen, with some of the electrons being scattered by the atoms of the specimen. An image, formed by a further electromagnetic objective lens, is projected onto a screen. Typically these electrons cause the screen material to fluoresce or phosphoresce, and the light is subsequently detected by a CCD camera.

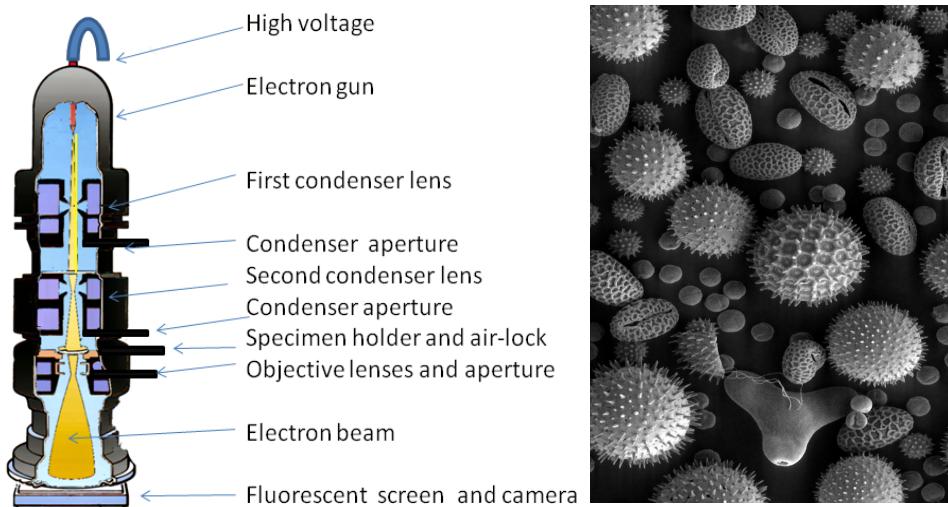


Figure 5.2: Schematic overview of the electron microscope. The picture on the right was taken using an electron microscope. The objects in the picture are pollen, which have a diameter of order $\sim 10 \mu\text{m}$ objects. The resolution which is obtained from the electron microscope image is much better than could be achieved using light. (Pictures taken from Wikipedia).

The diffraction limit of an electron microscope of this sort is tiny, because the de Broglie wavelength $\lambda \sim h/\sqrt{2m_e 100\text{keV}/e} = 3.9 \text{ pm}$! However we should remember that this is not the only restriction on imaging resolution. The imperfection of the focusing and objective electromagnetic lenses leads to *aberrations*, which in this case limit typical resolutions (for state of the art instruments) to 50 pm. This is still extremely impressive, allowing the positions of atoms on surfaces to be pinpointed.

Why use electrons rather than other types of particles. One advantage of electrons is that they are charged and therefore interact strongly with electric fields. This makes it possible to produce lenses by the use of inhomogeneous electric and magnetic fields (why must these be inhomogeneous? - return to the definition of a lens if you cannot answer this).

Neutron scattering

If we want to probe the internal structure of materials, electrons are not always the best choice. This is because they have a high cross-section for scattering. The methods of choice are therefore either to use X-ray scattering (which we saw before) or to use neutrons. Very often materials are examined using both of these methods. One important facility which does this near Zurich is the Paul-Scherrer institute, where they have both a neutron source and a high-intensity X-ray source.

Neutrons are produced in reactions in which the nuclei of atoms change. Typically this leads to the production of high-energy neutrons and a correspondingly small de Broglie wavelength

$$\lambda = \frac{h}{\sqrt{2m_n E_{KE}}} . \quad (5.5)$$

Neutrons principally interact with the nuclei in a material. In a similar manner to that used with X-rays material structure can be probed via Bragg-scattering. It is therefore preferable for the neutrons to have wavelengths of a similar size to the separation of atoms in the sample, which is on the order of an angstrom. In order to achieve this, the high-energy neutrons from the source are slowed down by passing them through a 'moderator' material. The energies required to give a de-Broglie wavelength of an angstrom are of order

$$E_{KE} = \frac{1}{2m_n} \frac{h^2}{(1\text{\AA})^2} = 0.08 \text{ eV} . \quad (5.6)$$

5.2 What is a matter wave?

5.2.1 The particle as a wavepacket

Let's return to the dynamics of matter waves. For matter to behave like a wave with momentum vector $p = \hbar/\lambda = \hbar k$, it must behave according to an equation for a wave in the same way that light does, namely that there is some sort of field with an amplitude which is a solution of the wave equation

$$A(x, t) = A \cos(kx - \omega t) = \frac{A}{2} \left(e^{i(kx - \omega t)} + e^{-i(kx - \omega t)} \right) . \quad (5.7)$$

Consider this wave in space - it extends throughout the whole of space. In what way can we expect to call this the equation of a particle then? Clearly a particle is localized in space, which surely means that it should not extend infinitely. In optics, we didn't worry too much about this, because at the time we covered optics we only thought in terms of waves, and in terms of intensity. The problem remains.

Instead of considering a wave with just one frequency and wavevector, let us consider summing over lots of waves which have the same amplitude but have wavevectors in the range $k - \Delta k \rightarrow k + \Delta k$. If all of these waves interfere, then the amplitude for the total wave can be written as

$$\psi(x, t) = \int_{k_0 - \Delta k}^{k_0 + \Delta k} A e^{i(kx - \omega(k)t)} dk \quad (5.8)$$

where $\omega(k)$ is the angular frequency associated with the wave with wavenumber k . Note that this looks very similar to the Fourier transforms which we encountered in optics. If we complete the analogy to light, we can associate the energy of each wave E with its frequency via $E = \hbar\omega$. However we should consider that we know the energy of a free particle - it is given by $E = p^2/2m$. Combining this with the de Broglie relation $p = \hbar k$ we obtain $\omega = \hbar k^2/2m$. If Δk is small, we can expand ω about the center frequency ω_0 as

$$\omega = \omega_0 + \frac{\partial \omega}{\partial k} (k - k_0) = \omega_0 + \frac{\hbar k_0}{m} (k - k_0) . \quad (5.9)$$

Substituting this result into our expression for $\psi(x, t)$, we find

$$\begin{aligned} \psi(x, t) &= A e^{i(\hbar k_0^2 t/m - \omega_0 t)} \int_{k_0 - \Delta k}^{k_0 + \Delta k} e^{ik(x - \hbar k_0 t/m)} dk \\ &= A e^{i(k_0 x - \omega_0 t)} \frac{2 \sin(\Delta k(x - \hbar k_0 t/m))}{(x - \hbar k_0 t/m)} . \end{aligned} \quad (5.10)$$

It is simpler to plot the real part of this expression

$$\Re(\psi(x, t)) = A \cos(k_0 x - \omega_0 t) \frac{2 \sin(\Delta k(x - \hbar k_0 t/m))}{(x - \hbar k_0 t/m)} \quad (5.11)$$

which is shown in figure 5.3. The amplitude of the wave as a function of x is given by a sinc function centered at $x = \hbar k_0 t/m$. Thus the position of the maximum changes in time according to

$$dx/dt = \hbar k_0 / m = p/m = v \quad (5.12)$$

We thus see that the wavepacket moves according to the same velocity as we would expect for a particle. The *group velocity* $d\omega/dk$ is the same as the particle velocity.

The second feature of equation 5.11 which should be noted is that for the particle to be localized to a smaller region in x , we require a larger range of k , ie. increased Δk . Since $\Delta p = \hbar \Delta k$, this requires Δp to increase. This is the feature that leads to the famed Heisenberg uncertainty relation (which we will come back to later in the lecture).

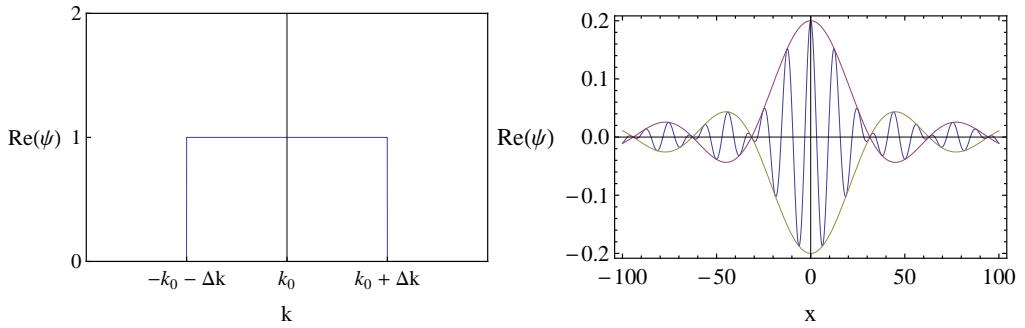


Figure 5.3: Real part of the wavepacket described by equation 5.11 plotted both in k and real space.

5.2.2 Interpretation of matter waves

We saw in the previous section that a non-monochromatic wave can produce amplitude localized at a particular point in space, and propagating with a group velocity which equals the particle velocity. However we have not said what this wave is. Think about the propagation of an electron wave. The electron has charge - are we really saying that this charge gets smeared out across space?

We are not. The waves we are starting to talk about here do not involve the smearing out of physical properties which we associate with particles. What we have here (and also in optics, in the modern interpretation) are waves with an amplitude which is related to a probability of finding a particle at a particular point *if we were to look for it there*. Consider a wave representing an electron given by the complex function $\psi(x, t)$. We say that the probability of finding the electron between position x and $x + dx$ is proportional to the *intensity* of the wave.

$$P(x, t)dx = |\psi(x, t)|^2 dx \quad (5.13)$$

or, in three dimensions

$$P(x, y, z, t)dxdydz = |\psi(x, y, z, t)|^2 dxdydz \quad (5.14)$$

What is “finding” the electron? What it means is that we get some signal which is local in space. For instance, if we look for diffraction with electrons, we might use as a detector some screen for which light is emitted from a certain point when an electron hits it. If we run the experiment with a single electron, we will get a single flash, from a single point. A diffraction pattern builds up because we do this for *many repeats which have the same probability distribution*, and thus build up the probability distribution, which displays the effects of interference. The electron does not interfere with other electrons, it interferes with itself.

It is important to note that if we are really going to use ψ to form a probability distribution, we must ensure that the probabilities add to 1 (the particle must be found somewhere in space). This means that

$$\iiint |\psi(x, y, z, t)|^2 dxdydz = 1 . \quad (5.15)$$

This is known as the *normalisation condition* for $\psi(x, y, z, t)$.

5.2.3 Diffraction and the uncertainty relation

With the discussion of the previous section in mind, let us revisit diffraction. In optics, we considered light incident on a screen with a slit in it. The angular distribution of the intensity of the light after the slit is given by a sinc function, such as that described by equation 5.3. For a beam falling straight onto a slit the position of the first minima was given for small angles by $\theta_1 = \lambda/\Delta_x$.

Now let us revisit this result in the context of particles. If the intensity is distributed over a range of angles, then there is a probability distribution for the direction of the momenta of the particles after the slit. Here we can view the slit as constraining the particle to pass through a restricted region of space. For all particles which pass through the screen, we know that they passed through a region of width Δ_x . Thus the uncertainty on position is given by $\Delta x = \Delta_x$. The probability distribution describing the range of angles which the particles have corresponds to uncertainty in the x -momentum of the particle p_x . Writing this uncertainty as Δp_x , we find that

$$\theta_1 = \frac{\Delta p_x}{p} \quad (5.16)$$

where p is the initial momentum of the particles (see figure 5.4). Therefore we find that

$$\frac{\Delta p_x}{p} = \frac{\lambda}{\Delta x} \quad (5.17)$$

which combined with the de Broglie relation $p = h/\lambda$ gives

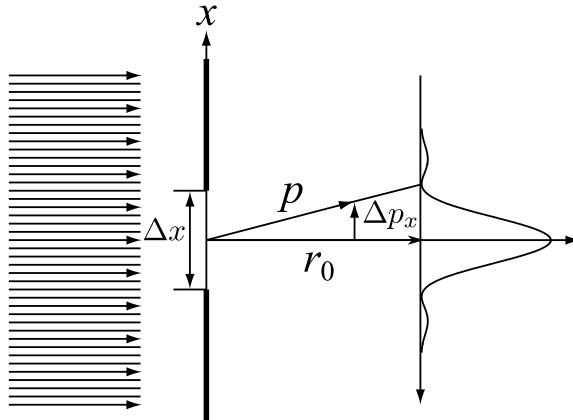


Figure 5.4: Diffraction of a wave from a slit, written in terms of the restrictions in position of a particle, and the resulting spread in momentum.

$$\Delta p_x \Delta x = h \quad . \quad (5.18)$$

This is a very rough derivation, but it gives a result which is striking. If we try and constrain a particle in space, as we do when we measure position to within a given accuracy, we must accept that the momentum will then be uncertain. The more precise statement of this condition is given by the renowned *Heisenberg uncertainty relation*

$$\Delta p_x \Delta x \geq \hbar \quad (5.19)$$

where $\hbar = h/(2\pi)$. This hints at a deeper property of quantum mechanics which is at odds with classical physics: we can not measure these two physical properties simultaneously with

arbitrary precision. Precise measurement of x will destroy all information about p_x , and vice-versa.

Such uncertainty relations do not only apply to x and p , but apply to a whole bunch of physical properties which can be measured (we call these observables). Another important relation relates energy and time

$$\Delta E \Delta t \geq \hbar. \quad (5.20)$$

At this point it is worth pointing out that the diffraction experiment is the same one carried out for classical light. It is in the interpretation that we attach some quantum mechanical meaning and derive an uncertainty relation. The important point here is that the particle will behave according to the probability distribution even if it is on its own - we will see this beautifully illustrated by the experiment described in the next section.

5.2.4 Double-slit interference with matter waves

One of the classic interference experiments is to look at the angular distribution of light after it passes through two slits separated by a small distance d . The emerging interference pattern is a clear visualization of a phenomena that can be attributed to waves. After our treatment of de Broglie waves you should not be surprised, that the same experiment can be run with matter waves as well. This gives a beautiful realization of the wave-particle nature of quantum mechanics.

Figure 5.5 shows the results of a two-slit interference experiment carried out using Helium atoms. The flux of particles is weak enough that particles arrive one at a time at the screen. Thus for small times (5 s or 50 s) we see an array of dots on the screen. However once a large number of particles have arrived at the screen it becomes clear that the particles are arriving according to a probability distribution which shows interference fringes. The interpretation is that the particles interfere with themselves. Thus we see a set of particles exhibiting interference which we associate with wave effects.

Such kinds of experiments have been carried out with a range of particles, from single photons to buckminsterfullerene molecules composed of hundreds of carbon atoms (these are also known as “bucky”balls).

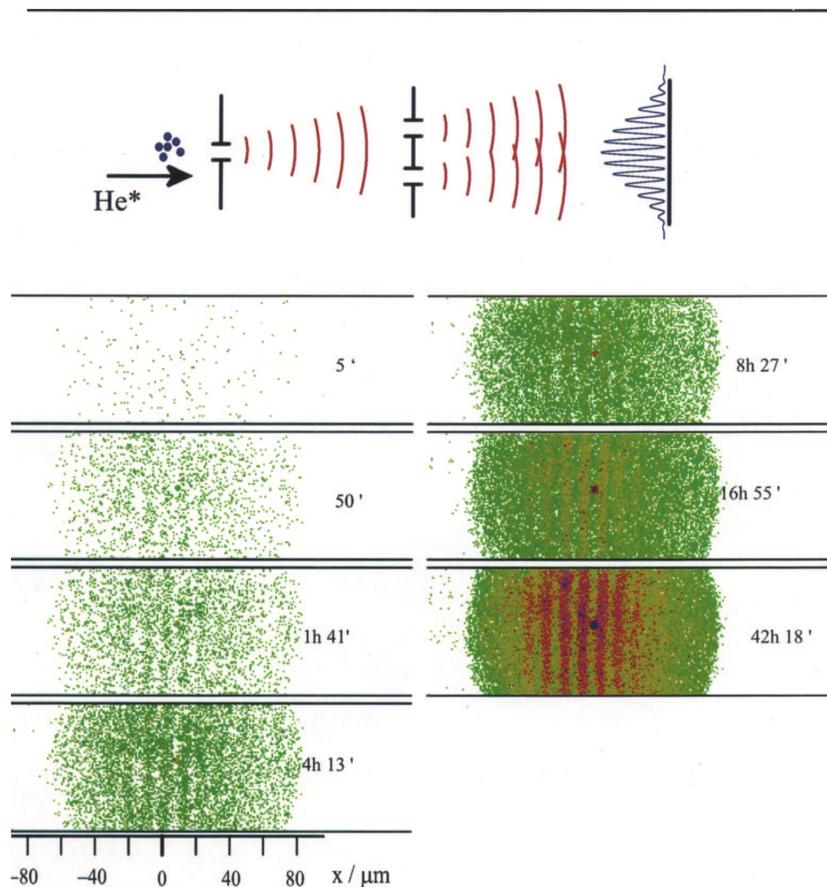


Figure 5.5: Helium atoms in an excited state are emitted from a small aperture in an oven. This forms a beam which passes a screen with two slits in it. The atoms then arrive at a screen, where the collision means that they are de-excited, and release energy which can be detected. The position at which they arrive is recorded as a function of time. Images are shown for times between 5 minutes (written as 5') and 42 hours and 18 minutes. (picture taken from the book by Haken and Wolf.)

Chapter 6

Quantum Mechanics

||||| HEAD We will now introduce quantum mechanics in a more formal manner. In the last few lectures we have become familiar with many of the features of systems at microscopic scales, including those which are now known to be the signatures of quantum mechanics. These include both particle and wave behavior being observed in a range of physical systems, and the agreement between treatments of blackbody radiation involving discrete energy levels and experimental data, rather than continua. These results pointed to the need for a completely new way of doing physics. Quantum theory provided this, and was based on several postulates. The neatest formulation of quantum mechanics is somewhat more mathematical than we want to go into in this course (you'll be more formally introduced in the theory track of the physics course), but I will introduce all of the basic concepts below.

Again you should make sure that you are reading a range of sources. While the notes should contain much of the basic information which is in the lectures, it is always useful to have a range of descriptions of complex topics. Each of us has been trained to think in our own independent way, which means that different explanations are understandable to different people. So for a conceptually complex topic such as quantum mechanics, it is well worth having a range of books. In my opinion this part of the course is covered at the appropriate level by "Physics of Atoms and Molecules" by B. H. Branden and C. J. Joachain, and in "The Physics of Atoms and Quanta" by H. Haken and H. C. Wolf. I think that the former does a better job.

Before starting on quantum theory, it is worth stepping back for a moment and asking what a physics theory which pertains to describe the fundamental workings of the world should contain. Think about classical mechanics. This provides a framework which has the following features

1. It tells us what parameters are needed to specify a particle. Some of these are fixed, such as the mass and charge. Others can change, such as the momentum and position.
2. It tells us how the properties of a particle evolves in time, even when subject to external potentials. This is what Newton's laws tell us.

In classical physics, knowledge of the position and momentum of a particle combined with knowledge of the external potentials allow us to predict exactly the behavior of the particle for all future times.

Our first requirement on quantum theory will therefore be that it provides us with a way of describing the object at one instance of time, and that it will tell us how this description

should evolve in time subject to external potentials. In addition, we should remember from our discussions of wave-particle duality that the type of measurement which we make on a system determines the result which we get, and how the system behaves after measurement. This should make us think that measurement in quantum physics is not very intuitive, and therefore our theory should provide the “rules” for measurement too.

6.1 The Wavefunction

- Every quantum system is completely described by a complex-valued function known as the wavefunction.

Let me emphasize *completely described*, all the information that specifies the system is contained in this function. For a wavefunction $\psi(x, t)$ describing a particle, the probability for finding the particle between x and $x+dx$ is $\psi^*(x, t)\psi(x, t)dx = |\psi(x, t)|^2dx$. The product $\psi^*(x, t)\psi(x, t)$ is therefore the probability density.

We know that probabilities must sum to one. Therefore the sum of the probabilities for finding the particle at any position must also equal one. Mathematically, this tells us that

$$\int_{-\infty}^{\infty} \psi^*(x, t)\psi(x, t)dx = 1 . \quad (6.1)$$

Despite my stating that all the information about the particle is contained in the wavefunction, it is not possible to extract this information in a measurement. The wavefunction does tell us the probabilities with which we get different measurement results.

We actually already used a wavefunction in the last lecture, though I tried to avoid calling it that. The wavefunction of a free particle with a momentum $p = \hbar k$ is

$$\psi(x, t) = Ae^{i(kx - \omega t)} \quad (6.2)$$

where $\hbar\omega = p^2/(2m)$. The complex constant A is added for normalization. As we learnt in the last lecture, if the momentum is well known, the wavefunction extends over all of space. The issue of normalization for the plane wave is somewhat tricky, so we will leave it alone at this point.

6.1.1 Expansion of functions

In quantum mechanics, it is very common to have to write the wavefunction in terms of an expansion over a set of functions belonging to an orthonormal set. Since we will find ourselves doing this a lot, it is worth me pointing out that you've done this sort of thing before, in the context of Fourier series expansions of waves, and indeed that the Fourier transform is another example of this type. Let's just remind ourselves what are the features of these methods.

Fourier series expansions

In a Fourier series, we attempt to write a real function $\psi(x)$ which is defined over a region $\pi \leq x \leq \pi$ as a sum over fixed sine or cosine functions. One example might be the cosine series, which we can use to write an arbitrary function so long as it is even (ie. it takes the same value for positive and negative x). In the cosine transform we write $\psi(x)$ as

$$\psi(x) = \sum_n \psi_n \cos(nx) \quad (6.3)$$

with the ψ_n to be determined, but not dependent on x . The reason that I am using the same symbol ψ here is that all the information about the function will be specified by these coefficients. There is no real reason for preferring the function $\psi(x)$ to the set of coefficients

ψ_n . This means that I should be as happy to use the set of coefficients ψ_n to write the function as writing $\psi(x)$ as a formula with x in.

In order to find the ψ_n we make use of the orthogonality of $\cos(nx)$ for different n , ie.

$$\int_{-\pi}^{\pi} \cos(mx) \cos(nx) dx = \pi \delta_{n,m} . \quad (6.4)$$

where $\delta_{n,m}$ is the Kronecker Delta Function, defined as

$$\begin{aligned} \delta_{n,m} &= 0 , \quad n \neq m \\ \delta_{n,m} &= 1 , \quad n = m . \end{aligned} \quad (6.5)$$

This allows us to multiply $\psi(x)$ through by $\cos(mx)$ and integrate over x (this was Fourier's great idea), which along with the orthogonality condition gives

$$\int_{-\pi}^{\pi} \cos(mx) \psi(x) dx = \int_{-\pi}^{\pi} \cos(mx) \sum_n \psi_n \cos(nx) dx \quad (6.6)$$

$$= \pi \sum_n \delta_{m,n} \psi_n = \pi \psi_m . \quad (6.7)$$

We then obtain a form for the co-efficients ψ_m :

$$\psi_m = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(mx) \psi(x) dx . \quad (6.8)$$

What features of the cosine functions did we use to obtain this form? Well, we used *orthogonality*, and we used the *normalization* of the functions. It is also worth noting that since all the functions we used for the expansion are cosines, we are restricted to expressing even functions. In order to express arbitrary functions we would be required to add terms involving sines to our series. Here we should note that sine is always orthogonal to cosine when integrated over the region $|x| \leq \pi$.

Expansions in terms of discrete sets of functions

Let us now look at a more general case. Let us assume we are given a set of complex functions $\phi_n(x)$ which we know are orthogonal, ie. they obey

$$\int \phi_n^*(x) \phi_m(x) dx = \delta_{m,n} \quad (6.9)$$

where we now take account of the complex nature of the functions and take the integral over the whole of x . We can then write the function $\psi(x)$ as

$$\psi(x) = \sum_n \psi_n \phi_n(x) . \quad (6.10)$$

Following the same procedure as we did for the Fourier cosine series we find

$$\int \phi_m^*(x) \psi(x) dx = \sum_n \psi_n \int \phi_m^*(x) \phi_n(x) dx \quad (6.11)$$

which combined with the orthogonality relation from equation 6.9 gives

$$\psi_m = \int \phi_m^*(x) \psi(x) dx . \quad (6.12)$$

It is worthwhile checking that these coefficients really do tell us the whole function. The way to do this is to insert the expression for ψ_m into equation 6.10. Here we should be careful, because the integral in the expression for ψ_m doesn't run over the same variable x we see in equation 6.10. Performing the re-insertion and differentiating between these two by the use of a x' and x , we find that

$$\psi(x) = \sum_m \int \phi_m^*(x') \psi(x') dx' \phi_m(x) \quad (6.13)$$

which we can re-order to be in the form

$$\psi(x) = \int \left(\sum_m \phi_m^*(x') \phi_m(x) \right) \psi(x') dx' . \quad (6.14)$$

For any of you who remember the Dirac delta function (those who don't will see it introduced in the next section), you should recognize that for this expression to give $\psi(x) = \psi(x)$, we require that

$$\left(\sum_m \phi_m^*(x') \phi_m(x) \right) = \delta(x - x') \quad (6.15)$$

where $\delta(x - x')$ is the Dirac delta function. This is known as a completeness relation - it tells us that we have a complete set of functions suitable for expanding any arbitrary function. For the purposes of our course there is no need to worry about completeness - all the sets of functions which we will use can safely be assumed to be complete.

Dirac delta function

There are many definitions of the Dirac delta function. The one which I find most useful is

$$\delta(x) = \int_{-\infty}^{\infty} e^{2\pi i \epsilon x} d\epsilon . \quad (6.16)$$

where the ϵ used here is a dummy variable which only appears inside the integral. The Dirac delta function is useful for evaluating integrals, but not as a function in its own right. The point to note is that if an integral over x contains the Dirac delta function multiplied by any function $f(x)$ then

$$\int_{-\infty}^{\infty} f(x) \delta(x) dx = f(0) . \quad (6.17)$$

The interpretation is that the Dirac delta function only has a value at $x = 0$, and is zero elsewhere (it is like an extremely sharp spike with an infinite peak height and infinitesimal width). This is the identity which I recognized in deriving the completeness condition above (equation 6.14). The area under the Dirac delta function is 1. This can be seen by setting $f(x) = 1$, which gives

$$\int_{-\infty}^{\infty} \delta(x) dx = 1 . \quad (6.18)$$

It will be useful to us to know that

$$\delta(ax) = \frac{1}{|a|} \delta(x) . \quad (6.19)$$

This can be shown from the original definition

$$\delta(ax) = \int_{-\infty}^{\infty} e^{2\pi i \epsilon ax} d\epsilon \quad (6.20)$$

which on making a change of variable $\chi = a\epsilon$, $d\chi = ad\epsilon$ (and remembering that the limits of the integral also get modified by factor $1/a$) gives

$$\delta(ax) = \frac{1}{|a|} \int_{-\infty}^{\infty} e^{2\pi i \chi x} d\chi = \frac{1}{|a|} \delta(x) \quad (6.21)$$

which is the result which we were looking for. If you're worried by the modulus sign, the explanation is that for negative a , the limits of the integral would change sign as well as $d\chi$ having the opposite sign to $d\epsilon$.

Expansions in terms of continuous sets of functions

A similar treatment as used for the discrete set of functions is required for continuous sets. In this case the orthogonality requirement is slightly different. Again we have a set of eigenfunctions, but this is a continuous set, labeled by a continuous variable α . We write the set of functions as $\phi(\alpha, x)$, and

$$\psi(x) = \int \tilde{\psi}(\alpha) \phi(\alpha, x) d\alpha \quad . \quad (6.22)$$

The orthogonality condition in this case requires a Dirac delta function

$$\int \phi^*(\alpha, x) \phi(\beta, x) dx = \delta(\alpha - \beta) \quad . \quad (6.23)$$

If we repeat the treatment performed above again here, we find that

$$\tilde{\psi}(\alpha) = \int \phi^*(\alpha, x) \psi(x) dx \quad . \quad (6.24)$$

The completeness relation for this case is similar to that we found before, but is now an integral

$$\int_{-\infty}^{\infty} \phi^*(\alpha, x') \phi(\alpha, x) dx' = \delta(x - x') \quad . \quad (6.25)$$

It is very common in quantum mechanics to make these expansions. As we will see in the next section, most of the things that you think of as variables, ie. energy, momentum, position have full sets of orthogonal functions associated with them (we call this a complete orthonormal set). So long as we can find these functions, and normalize them, we can write any wavefunction as a sum over a set of these functions. Which set we choose depends on the question that we want to answer.

6.2 Physical variables as operators

Let us return briefly to the wave-function of a free particle with momentum $p = \hbar k$ and energy $E = \hbar\omega$, given in equation 6.2. Note that if I apply the mathematical operation $\partial/\partial x$ to this wavefunction, I find

$$\begin{aligned}\frac{\partial}{\partial x}\psi(x, t) &= A \frac{\partial}{\partial x} e^{i(kx - \omega t)} \\ &= ik e^{i(kx - \omega t)} \\ &= ik\psi(x, t) = \frac{ip}{\hbar}\psi(x, t) .\end{aligned}\quad (6.26)$$

This means that this $\psi(x, t)$ satisfies an *eigenvalue* equation for the differential operator $\partial/\partial x$ with eigenvalue ip/\hbar . $\psi(x, t)$ is an *eigenstate* of this operator. The operator which would have $\psi(x, t)$ as an eigenfunction with eigenvalue p would be

$$-i\hbar \frac{\partial}{\partial x} . \quad (6.27)$$

We could perform the same analysis for the operator $\partial/\partial t$. There we find

$$\frac{\partial}{\partial t}\psi(x, t) = -i\omega\psi(x, t) = \frac{-iE}{\hbar}\psi(x, t) . \quad (6.28)$$

The operator which would have $\psi(x, t)$ as an eigenfunction with eigenvalue E would be

$$i\hbar \frac{\partial}{\partial t} . \quad (6.29)$$

It is one of the postulates of quantum mechanics that *these operators can always represent the momentum and energy* (not just for the case of free particles). The postulate of quantum mechanics is usually phrased as

- *Each physical observable is represented by an operator and the values obtained from measurements of the physical observable will be the eigenvalues of that operator.*

An example is that the position is given by the operator $\hat{x} = x$. It is easy to see that $\psi(x)$ is an eigenfunction of the position operator equation

$$\hat{x}\psi(x) = x\psi(x) . \quad (6.30)$$

I will use a hat whenever I want to write an operator. Written as above, the x example is particularly uninstructive. Consider instead the operator for momentum, which we take from our work on the free particle and write in terms of the position variable x as

$$\hat{p} = -i\hbar \frac{\partial}{\partial x} \quad (6.31)$$

or energy, which we can write in terms of time as

$$\hat{E} = i\hbar \frac{\partial}{\partial t} . \quad (6.32)$$

It is very important that since operators have to produce eigenvalues which correspond to physical variables such as momentum, position and energy, they should have real eigenvalues.

6.2.1 Change of basis: position and momentum

How can we use these operators? For instance, instead of asking the question “where is the particle?”, we might want to ask the question “what momentum does it have?”. The value we get would be an eigenvalue of the momentum operator. In order to obtain a probability distribution for these measurement results, it is very helpful to write the wave function in terms of the eigenstates of the momentum operator \hat{p} .

We start by finding a set of eigenfunctions for the momentum operator - these will provide a complete set of functions (playing a role analogous to that of the cosines in the Fourier series). The eigenfunction for which applying the momentum operator gives the eigenvalue p satisfies the equation

$$\hat{p}\phi(p, x) = p\phi(p, x) \quad (6.33)$$

$$-i\hbar \frac{\partial \phi(p, x)}{\partial x} = p\phi(p, x) . \quad (6.34)$$

$\phi(p, x)$ is the eigenfunction with eigenvalue p written as a function of x ,

$$\phi(p, x) = C(p)e^{ipx/\hbar} , \quad (6.35)$$

where $C(p)$ is a constant with respect to the integration variable x but may have a dependence on p . In order to use the expansions as we did for Fourier co-efficients we must check whether these satisfy orthonormality according to equation 6.23. This requires that

$$\int C^*(p)C(p')e^{i(p'-p)x/\hbar}dx = \delta(p' - p) . \quad (6.36)$$

Evaluating the left hand side results in

$$\int C^*(p)C(p')e^{i(p'-p)x/\hbar}dx = C^*(p)C(p') 2\pi\hbar\delta(p' - p) \quad (6.37)$$

and we therefore require

$$C^*(p)C(p') 2\pi\hbar\delta(p' - p) = \delta(p' - p) \quad (6.38)$$

We can find $C(p)$ by making use of the property of the Dirac delta function. Integrating both sides of this equation over p' we find that

$$\int C^*(p)C(p') 2\pi\hbar\delta(p' - p)dp' = 2\pi\hbar|C(p)|^2 = 1 \quad (6.39)$$

and we therefore obtain

$$|C(p)| = \frac{1}{\sqrt{2\pi\hbar}} . \quad (6.40)$$

Since we now have a set of states which have the correct properties, we can use equation 6.24 to find the wavefunction as a function of p

$$\tilde{\psi}(p) = \int \phi^*(p, x)\psi(x)dx . \quad (6.41)$$

In the case of position and momentum, we can see that since $\phi(p, x) = e^{ipx/\hbar}/\sqrt{2\pi\hbar}$ the momentum representation of the wavefunction is just the Fourier transform of the position representation

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{-ipx/\hbar}\psi(x)dx \quad (6.42)$$

with $p = \hbar k$. Correspondingly, to get back to the x representation given $\psi(p)$, we can invert the Fourier transform such that

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{ipx/\hbar} \psi(p) dp \quad (6.43)$$

Now that we have expressed the wavefunction in terms of the momentum, to predict what probability we have of finding a value between p and $p + dp$ on measuring the momentum, we take

$$\tilde{\psi}^*(p)\tilde{\psi}(p)dp = |\tilde{\psi}(p)|^2 dp \quad (6.44)$$

just as we did for the position.

Example: Imagine that I gave you the wave function, written in terms of the spatial co-ordinate x

$$\psi(x) = N(e^{ik_1 x} + e^{-ik_1 x}) \quad (6.45)$$

and I asked what the momentum distribution is? You would now start by finding the distribution in the p representation

$$\tilde{\psi}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} e^{-ipx/\hbar} N(e^{ik_1 x} + e^{-ik_1 x}) dx \quad (6.46)$$

$$= N \sqrt{\frac{2\pi}{\hbar}} (\delta(p/\hbar + k_1) + \delta(p/\hbar - k_1)) \quad (6.47)$$

where again I am using the the Dirac delta function. The distribution therefore consists of two sharp spikes at $p = \hbar k_1$ and $p = -\hbar k_1$. The amplitude of each of the delta functions is the same, so I will have an equal probability of finding the particle with each of these two momenta.

6.3 Measurement

So far, I have phrased the description of quantum mechanics only in terms of position and momentum. This gave some idea that the wavefunction can predict probabilities for obtaining different measurement results. I now want to be more precise though, so let us clarify what the rules for measurement are in quantum mechanics.

In the previous section we expressed the wavefunction as a continuous sum over momentum eigenstates. This is an example of the superposition principle in quantum mechanics. This states that if we add two valid quantum states together we will get another valid quantum state. The simplest superposition which I can think of is a superposition of two eigenstates of an operator \hat{O} , which have eigenvalues O_1 and O_2

$$\psi(x) = a_1\phi_1(x) + a_2\phi_2(x) . \quad (6.48)$$

where a_1 and a_2 are complex numbers (these have to satisfy a simple constraint, as we will see below).

The rule of measurement in quantum mechanics states that a

- A measurement made on $\psi(x)$ of the physical quantity corresponding to the operator \hat{O} will yield one of the two eigenvalues O_1 and O_2 . The probability to get O_1 is given by $|a_1|^2$ and the probability to get O_2 is given by $|a_2|^2$.
- After the measurement, the system will be in the eigenstate of \hat{O} which has eigenvalue corresponding to the measurement result.

Therefore if the value O_1 was obtained on measuring the system, then the state of the system after the measurement is given by $\phi_1(x)$. Once again we have to check that our probabilities add to equal 1. This requires that

$$|a_1|^2 + |a_2|^2 = 1 . \quad (6.49)$$

You should note that repeated measurement of the same variable will give the same results. If on the first measurement, we obtain the eigenvalue O_1 , then after the measurement the state of the system is $\phi_1(x)$. This is the eigenstate of the measurement with eigenvalue O_1 . A subsequent measurement of the same operator will thus also obtain the value O_1 .

These are the rules for measurement. Measurement is the strangest feature of quantum mechanics. On extracting information from a physical system, we completely change its state. A wavefunction might describe probability amplitudes for a particle to appear over a vast region of space, but as soon as we measure the position of a particle at some location we instantaneously collapse the whole thing to a point. You'd be right to find this totally bizarre - I think is the aspect of quantum mechanics which causes physicists the most discomfort too.

6.3.1 Example: position and momentum

We saw an example in the set of eigenfunctions for momentum where the momentum eigenstate with eigenvalue p is written in terms of x as

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}}e^{ipx/\hbar} . \quad (6.50)$$

and in terms of p as

$$\begin{aligned}\tilde{\psi}(p') &= \frac{1}{\sqrt{2\pi\hbar}} \int e^{-ip'x/\hbar} \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar} dx \\ &= \delta(p - p')\end{aligned}\quad (6.51)$$

If the momentum of a particle in this state is measured, the answer obtained would certainly be p because the delta function is very sharply peaked at this value.

If, on the other hand we chose to measure x , we would obtain a value between x' and $x' + dx$ with probability given by

$$\psi^*(x')\psi(x')dx \quad (6.52)$$

Let us say that we did obtain the value x' . The state of the system after the measurement would be a delta function in x

$$\psi_{\text{after}}(x) = \delta(x - x') \quad (6.53)$$

and thus subsequent position measurements would also give the value x' . However the momentum is now completely unknown

$$\tilde{\psi}_{\text{after}}(p) = \frac{1}{\sqrt{2\pi\hbar}} \int e^{ipx/\hbar} \delta(x - x') dx = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx'/\hbar} \quad (6.54)$$

and so the result of a subsequent measurement of the momentum could give result p' with probabilities corresponding to $|\psi_{\text{after}}(p')|^2 dp$.

6.3.2 Simultaneous sets of eigenstates

In physics generally, we try to find sets of properties which we can attach to a particle or another type of physical system. For classical physics, we might assign variables x, p, E to the position, momentum, energy of the particle. In classical physics, all of these could be well defined, in the sense that the particle can take all of these values simultaneously.

We know already, from our discussion of the wave nature of matter, that in quantum mechanics it is not possible to specify the position and momentum simultaneously with infinite precision. The best we can do is to specify the wave function for the particle (this gives us *all* information). However we have just seen that there are certain states for which one of these variables is completely determined - these are the eigenstates of the operator associated with that variable. What if we want to determine not only the x position, but also the y position - is this possible? Another example might be whether we can have states with well specified energy E and momentum p ?

In either of these cases, in order for the values of the measurement to be compatible we require that the system can simultaneously exist in an eigenstate of *both* operators. Is there a way to find this out without finding all of the eigenstates of each operator and comparing them?

There is. We can write the condition that operators \hat{O} and \hat{Q} share an eigenfunction mathematically as

$$\hat{O}\phi = O\phi \quad (6.55)$$

$$\hat{Q}\phi = Q\phi . \quad (6.56)$$

Let us now apply the operators $\hat{O}\hat{Q}$ and $\hat{Q}\hat{O}$ to ϕ (note that in each case here the *order* in which the operators are applied is exchanged). We find

$$\hat{O}\hat{Q}\phi = Q\hat{O}\phi = QO\phi \quad (6.57)$$

$$\hat{Q}\hat{O}\phi = O\hat{Q}\phi = OQ\phi = QO\phi \quad (6.58)$$

where on the second line the order of the eigenvalues can be reversed because they are numbers, not operators. Therefore a condition for ϕ to be a simultaneous eigenstate of both \hat{Q} and \hat{O} is that

$$(\hat{O}\hat{Q} - \hat{Q}\hat{O})\phi = 0 . \quad (6.59)$$

In many cases we want to find whole sets of eigenfunctions for which equations 6.55 and 6.56 are satisfied, because this would allow us to express an arbitrary wave function in terms of these functions. In this case we require that equation 6.59 is satisfied for all eigenstates ϕ . This then requires that the operator

$$\hat{O}\hat{Q} - \hat{Q}\hat{O} = 0 , \quad (6.60)$$

which means that *this operator gives the value zero when applied to any state*. We call the left hand side a commutator, and it is generally written in the form

$$[\hat{O}, \hat{Q}] \equiv \hat{O}\hat{Q} - \hat{Q}\hat{O} . \quad (6.61)$$

If two operators *commute*, then equation 6.60 is satisfied and they share a set of eigenstates. This means that the values of the two physical observables which are represented by these operators can be known with infinite precision simultaneously.

The alternative case is that equation 6.60 is not satisfied. In this case the converse is true, and the two physical observables cannot be obtained to infinite precision. This is the case for position and momentum and lies at the core of the Heisenberg uncertainty relation.

Example 1: Evaluating the commutator for \hat{x} and \hat{p} , we find

$$\left[x, -i\hbar \frac{\partial}{\partial x} \right] \psi = x \left(-i\hbar \frac{\partial \psi}{\partial x} \right) + i\hbar \frac{\partial}{\partial x} (x\psi) \quad (6.62)$$

$$= -i\hbar x \frac{\partial \psi}{\partial x} + i\hbar \psi + i\hbar x \frac{\partial \psi}{\partial x} \quad (6.63)$$

$$= i\hbar \psi \quad (6.64)$$

so the commutator does not equal zero. Instead we find that

$$[\hat{x}, \hat{p}] = i\hbar . \quad (6.65)$$

This means that \hat{x} and \hat{p} do not share a set of eigenfunctions, and thus cannot have definite values simultaneously. One consequence of this is the Heisenberg uncertainty principle.

Example 2: Can we have simultaneous eigenstates of momentum and kinetic energy? The kinetic energy operator is given by $p^2/(2m)$. Therefore we take

$$\left[\frac{\hat{p}^2}{2m}, \hat{p} \right] \psi = \frac{1}{2m} [\hat{p}^2, \hat{p}] \psi = \frac{1}{2m} (\hat{p}^3 - \hat{p}^3) \psi = 0 . \quad (6.66)$$

An operator always commutes with itself. So yes, we can have simultaneous eigenstates of kinetic energy and momentum - de Broglie waves are these solutions.

Other examples which are useful to note are

$$[\hat{x}, \hat{y}] = 0 \quad (6.67)$$

$$[\hat{x}, \hat{p}_y] = 0 \quad (6.68)$$

where $\hat{p}_y = -i\hbar\partial/\partial y$ is the operator for the y component of the momentum.

6.3.3 Expectation values

As opposed to finding the probabilities to find the position of the particle between x and $x + dx$, another question we might ask is “what is the average value of the position?”. This is not a quantity which we can obtain from a single particle. It could only be obtained by taking many particles which are prepared in exactly the same way (and thus have identical wavefunctions), and measuring the same physical variable on all of them.

For a physical variable represented by an operator \hat{O} , we find the expectation value by taking

$$\langle \hat{O} \rangle = \int \psi^*(x) \hat{O} \psi(x) dx \quad (6.69)$$

where the integral is again taken over all x .

An example would be the expectation value for the kinetic energy, for which $\hat{O} = \hat{p}^2/(2m)$, which is

$$\langle E_{\text{KE}} \rangle = \frac{-\hbar^2}{2m} \int \psi^*(x) \frac{\partial^2}{\partial x^2} \psi(x) dx \quad (6.70)$$

where equation 6.31 has been substituted for the momentum operator. Note that the operator appears to the right of the complex conjugate of the wave function. This is important, because our differential operator acts to the right; if we put the operator in the wrong place we would get a different (and wrong) result.

If the particle is in an eigenstate of an operator, then the expectation value is also completely determined: it is the eigenvalue of the eigenstate. This follows from the eigenvalue equation

$$\hat{O}\phi_O = O\phi_O \quad (6.71)$$

which results in

$$\langle \hat{O} \rangle = \int \phi_O^* \hat{O} \phi_O dx = O \int \phi_O^* \phi_O dx = O . \quad (6.72)$$

As an example of the evaluation of an expectation value, let us try to evaluate the expectation value of energy for a superposition of two energy eigenstates $\psi_1(x)$ and $\psi_2(x)$, which have energy E_1 and E_2 respectively. The superposition is given by

$$\psi = a_1\psi_1 + a_2\psi_2 \quad (6.73)$$

where I have dropped the dependence on x to simplify the notation. The expectation value for energy is

$$\begin{aligned} \langle \hat{E} \rangle &= \int \left[(a_1^* \psi_1^* + a_2^* \psi_2^*) \hat{E} (a_1 \psi_1 + a_2 \psi_2) \right] dx \\ &= \int [(a_1^* \psi_1^* + a_2^* \psi_2^*) (a_1 E_1 \psi_1 + a_2 E_2 \psi_2)] dx \\ &= |a_1|^2 E_1 \int \psi_1^* \psi_1 dx + |a_2|^2 E_2 \int \psi_2^* \psi_2 dx \end{aligned} \quad (6.74)$$

where we used the orthogonality relation between the eigenstates of \hat{E} to find that the cross terms equal zero. The normalization of the eigenfunctions allows us to evaluate both integrals to 1. Since the values of $|a_1|^2$ and $|a_2|^2$ are the probabilities p_1 and p_2 to find the system at each energy, we find that

$$\langle \hat{E} \rangle = p_1 E_1 + p_2 E_2 \quad (6.75)$$

which is what we would expect from our studies of probability distributions earlier in the course.

6.4 The Schrödinger equation

In the previous section, we found how to relate position to momentum, and energy and time. However we showed no link between these two. You might be thinking that it is a bit strange that the momentum operator doesn't seem to depend on time - it is related to the velocity through $p = mdx/dt$ isn't it?

The final postulate of quantum mechanics tells us how wave functions evolve in time

- The wave function $\psi(x, y, z, t)$ evolves in time according to the Schrödinger equation

$$\hat{H}\psi(x, y, z, t) = i\hbar \frac{\partial\psi(x, y, z, t)}{\partial t} \quad (6.76)$$

where \hat{H} is called the Hamiltonian operator.

The Hamiltonian is therefore a very important operator, because it governs the time evolution of the system. It is the sum over operators for kinetic and potential energy. For a single particle of mass m in a static potential $V(x, y, z)$ it has the form

$$\frac{1}{2m} (\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2) + V(x, y, z) \quad (6.77)$$

which using the differential form for the momentum operators gives

$$\hat{H} = \frac{-\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V(x, y, z) \quad (6.78)$$

The eigenstates of the Hamiltonian are the states with a given energy - in this sense the Hamiltonian offers another form of the operator for the energy. The eigenvalues of the Hamiltonian are the energies which the system can take. These eigenfunctions satisfy the Time Independent Schrödinger Equation (TISE)

$$\hat{H}\phi_E = E\phi_E \quad (6.79)$$

The eigenstates of energy are of great importance in quantum mechanics, because they evolve in time in a very predictable manner. We can see this from

$$\hat{H}\phi_E = E\phi_E = i\hbar \frac{\partial\phi_E}{\partial t} \quad (6.80)$$

which has solutions

$$\phi_E(x, y, z, t) = e^{-iEt/\hbar} \phi_E(x, y, z, 0) \quad (6.81)$$

In other words, an eigenstate of energy evolves in time by acquiring a phase $e^{-iEt/\hbar}$ where E is the energy eigenvalue. The probability to find the particle between $\mathbf{r} \equiv (x, y, z)$ and $\mathbf{r} + d\mathbf{r} \equiv (x + dx, y + dy, z + dz)$ is given by

$$\begin{aligned} \phi_E^*(\mathbf{r}, t)\phi_E(\mathbf{r}, t)d\mathbf{r} &= e^{iEt/\hbar} \phi_E^*(\mathbf{r}, 0)e^{-iEt/\hbar} \phi_E(\mathbf{r}, 0)d\mathbf{r} \\ &= \phi_E^*(\mathbf{r}, 0)\phi_E(\mathbf{r}, 0)d\mathbf{r} \end{aligned} \quad (6.82)$$

ie. the probabilities for a particle in an eigenstate do not change in time.

We can see from this example that the energy eigenstates and eigenvalues are extremely important. If we can find them out, we can find the state of the system for all time. The key to finding these solutions lies in solving Schrödinger's equation for the problem. Each problem varies in the form of the potential $V(x, y, z)$, and in the necessary boundary conditions which must be satisfied by the wave function.

For the next few lectures we will look at a variety of situations, obtain the Hamiltonians, and derive the solutions to the TISE. This will allow us to explore simple features of wave functions, and will also allow us to find the wavefunctions and energy levels of the Hydrogen atom. The limitations of the methods I will describe occur when the Hamiltonian itself becomes time dependent. Time dependence of the Hamiltonian results in dynamics which is generally much more complex.

6.4.1 The free particle in 1-D

The first solution to Schrödinger's equation which we'll look at is the free particle. This is relatively simple. The Hamiltonian for a free particle is just the kinetic energy operator, and thus the energy eigenfunctions $\psi_E(x)$ when written in terms of x satisfy

$$-\frac{\hbar^2}{2m} \frac{\partial \psi_E(x, 0)}{\partial x^2} = E\psi_E(x, 0) \quad (6.83)$$

which has solutions of the form

$$\psi_E(x) = Ae^{ikx} + Be^{-ikx} \quad (6.84)$$

where

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad (6.85)$$

and A and B are constants to be determined. The spectrum of energies which the particle can take is continuous.

We can also make use of the energy operator to find

$$i\hbar \frac{\partial \psi_E(x, t)}{\partial t} = E\psi_E(x, t) \quad (6.86)$$

and therefore we find that wavefunctions which solve the Time Dependent Schrödinger Equation (TDSE) with energy E are of the form

$$\psi_E(x, t) = e^{-i\omega t} (Ae^{ikx} + Be^{-ikx}) \quad (6.87)$$

with $\omega = E/\hbar$ and k as given above. These are the same as the plane wave solutions that we used in optics and also used in the context of de Broglie waves. Thus the wave function is exactly of the form of the amplitudes we used previously, and the results of measurements on a free particle will produce outcomes which are dependent on effects such as diffraction and interference.

Since the energy eigenstates of the Schrödinger equation are so important, we will spend the next few lectures finding these solutions for a number of simple potentials, and examining what features we find. In each case we will examine both the energy eigenfunctions and the energy eigenvalues. As with any other differential equation, it is not only the form of the equation which is of interest in defining the eigenfunctions, but also the boundary conditions.

6.5 The Infinite Square Well

The simplest potential function which can be treated in quantum mechanics is the 1-D infinite square well as shown in figure 6.1, for which

$$\begin{aligned} V(x) &= 0, & a \geq x \geq 0 \\ V(x) &= \infty, & x > a \quad x < 0 . \end{aligned} \quad (6.88)$$

Since the energy of a particle in such a potential will be infinite if it enters the region $x > a$ and $x < 0$, the boundary conditions are simple, that the wave function cannot penetrate into the walls of the well.

For the region $a \geq x \geq 0$, since the potential is zero, the Hamiltonian is

$$\hat{H} = \frac{\hat{p}^2}{2m} = \frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} . \quad (6.89)$$

The TISE for this problem is therefore

$$\frac{-\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} = E\psi(x) , \quad (6.90)$$

which can be re-arranged into a more familiar form

$$\frac{\partial^2 \psi(x)}{\partial x^2} = -\frac{2mE}{\hbar^2} \psi(x) . \quad (6.91)$$

We wish to find the solutions to this equation with positive energies, because the kinetic energy of a particle is always positive. These are sine or cosine functions, which we can write as

$$\psi(x) = A \sin(kx) + B \cos(kx) \quad (6.92)$$

where

$$k = \sqrt{\frac{2mE}{\hbar^2}} \quad (6.93)$$

and A and B are complex co-efficients.

The boundary condition $\psi(x = 0) = 0$, excludes all cosine solutions. In order for the sine solutions to vanish at the other boundary $x = a$, we require

$$A \sin(ka) = 0 \quad (6.94)$$

which is satisfied if $ka = n\pi$ with n an integer. This gives a restriction on the possible energy eigenvalues E_n

$$\sqrt{\frac{2mE_n}{\hbar^2}} = \frac{n\pi}{a} \quad (6.95)$$

leading to a discrete set of energy eigenvalues

$$E_n = \frac{n^2 \pi^2 \hbar^2}{2ma^2} \quad (6.96)$$

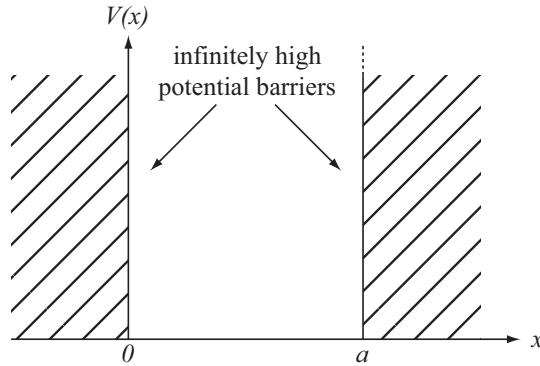


Figure 6.1: Potential with infinitely high walls at 0 and a .

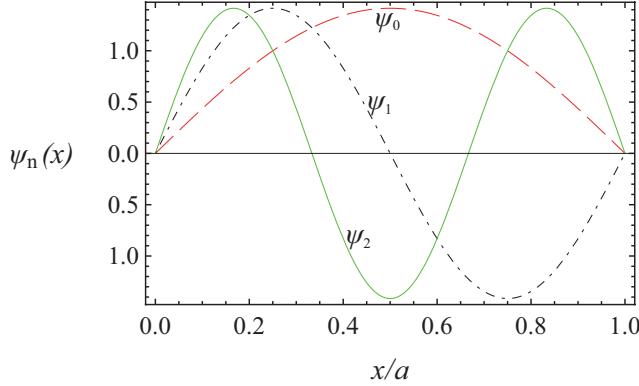


Figure 6.2: The three lowest eigenfunctions of energy for the infinite square well. Ground state: dashed, red. First excited state: dot-dashed, black. Second excited state: solid, green.

This is our first example of discrete energy levels, a feature of quantum mechanics which we expected from the earlier work we did on Planck's spectrum of blackbody radiation. It is important to note where the discrete nature of these energy levels comes from. We get the discrete levels because we have to satisfy boundary conditions - the wavefunction is confined to a small region. We will find that this result is more generally true. Whenever we have a bound state in quantum mechanics (one where the potential forces the particle to be in a particular region) we get discrete energy levels.

The corresponding eigenstates are

$$\psi_n(x) = A_n \sin(n\pi x/a), \quad a \geq x \geq 0 \quad (6.97)$$

where the normalization constant A_n remains to be determined. We do this by ensuring that

$$\int_{-\infty}^{\infty} \psi_n^*(x) \psi_n(x) dx = 1 , \quad (6.98)$$

for every eigenfunction. This gives

$$\begin{aligned} 1 &= \int_0^a |\psi_n(x)|^2 dx \\ &= |A_n|^2 \int_0^a \sin^2(n\pi x/a) dx \\ &= |A_n|^2 \frac{a}{2} . \end{aligned} \quad (6.99)$$

The requirement for normalization is thus that $A_n = \sqrt{2/a}$. Finally, the eigenfunctions are

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin(n\pi x/a) . \quad (6.100)$$

The first few eigenfunctions are plotted in figure 6.2. This provides us with a good opportunity to examine general features of wavefunctions.

First, note that the lowest energy eigenstate (we call this the ground state) does not cross zero. It does go to zero, at the boundaries of the potential well, but it never crosses zero. The first excited state, ie. the one with the next highest energy, crosses zero once, while the next excited state crosses twice. This is a general rule for eigenstates of the energy operator, which we will see again in later examples.

Second, note that the ground state has lower gradients and lower curvature. You should be able to see where this is coming from. The energy of the particle inside the well is purely given by the kinetic energy, resulting in a Hamiltonian operator which is related to the curvature of the wavefunction. Therefore it is obvious that a state with lower energy will have lower curvature. Later, we will see wave functions for which the curvature changes with position. This indicates that the kinetic energy of the particle also varies with position.

6.6 The Finite Square Well

Let us now consider a potential which takes a finite value V_0 for $|x| > a/2$ and is zero for $|x| \leq a/2$. This is the finite version of the previous potential, but displaced in space by $a/2$.

In the region inside the potential well, the potential is zero, and thus the Hamiltonian is the same as that considered previously. We should therefore find the same form of the solutions. However, we should also note that the boundary conditions have changed, because the wave function can now take finite values at the edge of the well (the energy of the particle inside the walls is no longer infinite). Since we can no longer assume that the cosine terms in equation 6.92 are zero, we must return to this general form. First though, let us investigate whether the general properties of the Hamiltonian can help us find the solutions.

6.6.1 Parity inversion

Prior to looking at the boundary conditions, let us first consider the potential. It is symmetric about the origin, therefore the energy does not change when we invert the co-ordinate system, ie. we make the replacement $x \rightarrow -x$. We can write an operator for inverting the co-ordinate system as

$$\hat{P}x = -x \quad (6.101)$$

which acts on a wavefunction written in terms of x to give

$$\hat{P}\psi(x) = \psi(-x) . \quad (6.102)$$

Operating twice on the wavefunction, we will regain our original wavefunction.

$$\hat{P}^2\psi(x) = \hat{P}\psi(-x) = \psi(x) . \quad (6.103)$$

Therefore the eigenvalue of \hat{P}^2 is 1, and the eigenfunctions for \hat{P}^2 include any wave function.

If the Hamiltonian commutes with the inversion operator \hat{P} , then the Hamiltonian will share a set of eigenstates with the inversion operator. Since our Hamiltonian is quadratic in the co-ordinate, it will commute. The next question to ask is what the possible eigenvalues of the inversion operator are? To see this, note that an eigenstate of \hat{P} satisfies the equation

$$\hat{P}\psi_P(x) = P\psi_P(x) = \psi_P(-x) . \quad (6.104)$$

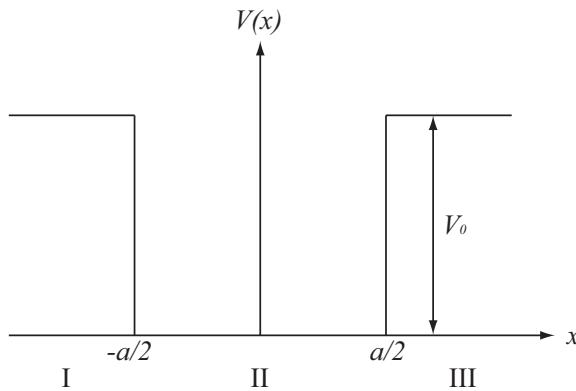


Figure 6.3: A potential well with walls of height V_0 .

where P is the parity eigenvalue. Operating for a second time on the eigenfunction, we find

$$\hat{P} \hat{P} \psi_P(x) = \hat{P} P \psi_P(x) = P^2 \psi_P(x) , \quad (6.105)$$

which means that $P^2 = 1$ and thus $P = \pm 1$. So if our Hamiltonian is symmetric with respect to parity inversion, we can expect that the eigenstates of energy are either symmetric ($P = +1$) or antisymmetric ($P = -1$) with respect to inversion.

6.6.2 Continuity of the wave function

We should remember that the momentum is given by the operator

$$\hat{p} = -i\hbar \frac{\partial}{\partial x} \quad (6.106)$$

and the kinetic energy is

$$E_{KE} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} . \quad (6.107)$$

If our particle has finite energy, we require that both the kinetic energy and momentum are also finite over the wave function. Ensuring that the momentum is finite forces the wave function to be continuous, since a discontinuous wave function has infinite gradient. Therefore we have the first set of conditions for any boundary at any position x_0

$$\lim_{dx \rightarrow 0} \psi(x_0 - dx) = \lim_{dx \rightarrow 0} \psi(x_0 + dx) \quad (6.108)$$

If we know that the energy is finite, the Schrödinger equation ensures that the gradient of the wave function must be continuous. Consider re-writing the Schrödinger equation at the position x_0 as

$$\begin{aligned} \left. \frac{\partial^2 \psi}{\partial x^2} \right|_{x_0} &= \frac{\left. \frac{\partial \psi}{\partial x} \right|_{x_0+dx/2} - \left. \frac{\partial \psi}{\partial x} \right|_{x_0-dx/2}}{dx} \\ &= -\frac{2m(E - V(x_0))}{\hbar^2} \psi(x_0) \end{aligned} \quad (6.109)$$

Multiplying through by dx , we find that

$$\left. \frac{\partial \psi}{\partial x} \right|_{x_0+dx/2} - \left. \frac{\partial \psi}{\partial x} \right|_{x_0-dx/2} = -\frac{2m(E - V(x_0))}{\hbar^2} \psi(x_0) dx \quad (6.110)$$

which in the limit that $dx \rightarrow 0$ gives us

$$\left. \frac{\partial \psi}{\partial x} \right|_{x_0+dx/2} - \left. \frac{\partial \psi}{\partial x} \right|_{x_0-dx/2} = 0 \quad (6.111)$$

so long as the both the energy and the potential are finite.

Note that for the *infinite* square well the wave function was zero outside the well, and had zero gradient. There we ensured that the wave function was continuous, but since the potential is infinite, the gradient of the wavefunction was *not* continuous across the boundary. *This is because this is not a real physical situation, because of the infinite potential.* For any real physical situation, the derivatives with respect to x are continuous across the boundary.

6.6.3 Solutions of the TISE

Let us now go back to our original problem. Let us label three different regions. Region *I* corresponds to $x < -a/2$, region *II* to $|x| \leq a/2$ and region *III* to $x > a/2$. I shall therefore split the eigenfunction of the TISE into the three parts which exist in these three areas $u_I(x)$, $u_{II}(x)$, and $u_{III}(x)$.

Inside the well (region *II*), the TISE gives

$$\frac{\partial^2 u_{II}(x)}{\partial x^2} = -\frac{2mE}{\hbar^2} u_{II}(x) \quad (6.112)$$

with solutions

$$u_{II}(x) = A \cos(kx) + B \sin(kx) \quad (6.113)$$

and

$$k^2 = \frac{2mE}{\hbar^2} \quad (6.114)$$

as for the case of the infinite well.

In regions *I* and *III*, the TISE is

$$\frac{-\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} + V_0 \psi(x) = E \psi(x) , \quad (6.115)$$

which can be re-arranged to

$$\frac{\partial^2 \psi(x)}{\partial x^2} = -\frac{2m}{\hbar^2} (E - V_0) \psi(x) . \quad (6.116)$$

The form of the solutions will be different depending on whether $E - V_0 > 0$ or $E - V_0 < 0$. We therefore separate out these two cases below. In both cases, we make use of the symmetry of the wave function to simplify the problem.

Solutions for $E - V_0 < 0$

If the energy of the particle is smaller than the potential of the walls, then for $|x| > a/2$ we have solutions

$$u_I(x) = a_I e^{\kappa x} + b_I e^{-\kappa x} \text{ for } x < -a/2 \quad (6.117)$$

$$u_{III}(x) = a_{III} e^{\kappa x} + b_{III} e^{-\kappa x} \text{ for } x > a/2 \quad (6.118)$$

where

$$\kappa = \sqrt{\frac{2m}{\hbar^2} (V_0 - E)} \quad (6.119)$$

is real valued. For these solutions, one boundary condition is clear. As $|x| \rightarrow \infty$, we must ensure that the wave function does not expand exponentially, or else the energy will be infinite. We can therefore immediately set $b_I = 0$ and $a_{III} = 0$. The solutions then simplify to

$$u_I(x) = a_I e^{\kappa x} \text{ for } x < -a/2 \quad (6.120)$$

$$u_{III}(x) = b_{III} e^{-\kappa x} \text{ for } x > a/2 . \quad (6.121)$$

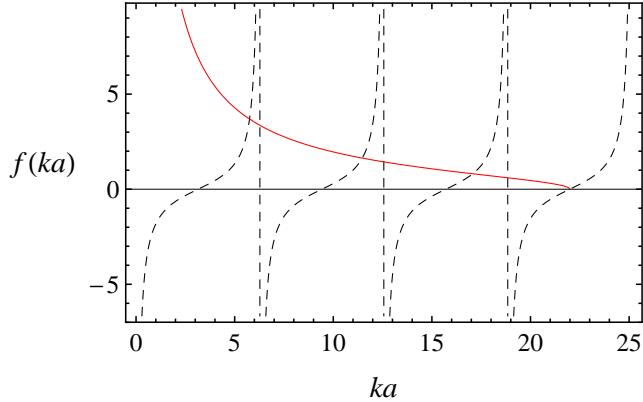


Figure 6.4: L.H.S. (dashed, black) and R.H.S. (solid, red) of equation 6.128 for $\chi a = 7\pi$.

First let us consider the boundary conditions for the *anti-symmetric* solution. In this case $u_{II}(x) = B \sin(kx)$ and $u_I(-x) = -u_{III}(x)$. We therefore know immediately that $a_I = -b_{III}$. There is therefore a simple relation between B , a_I and b_{III}

$$a_I e^{-\kappa a/2} = -b_{III} e^{-\kappa a/2} = -B \sin(ka/2) . \quad (6.122)$$

The solutions will therefore be

$$u_I(x) = -B \sin(ka/2) e^{\kappa(a/2+x)} \quad (6.123)$$

$$u_{II}(x) = B \sin(kx) \quad (6.124)$$

$$u_{III}(x) = B \sin(ka/2) e^{\kappa(a/2-x)} \quad (6.125)$$

Now let us check that the gradient of the wave function is continuous. This requires

$$\frac{\partial u_I}{\partial x} \Big|_{-a/2} = \frac{\partial u_{II}}{\partial x} \Big|_{-a/2} \Rightarrow -\kappa B \sin(ka/2) = kB \cos(ka/2) \quad (6.126)$$

The solutions of this equation define the allowed energy eigenvalues. Remember that κ is already related to k , through the definitions of these two variables. We can write κ in terms of k as

$$\kappa^2 = \chi^2 - k^2 \quad (6.127)$$

with $\chi^2 = \frac{2mV_0}{\hbar^2}$ resulting in

$$-\cot(ka/2) = \frac{\sqrt{\chi^2 - k^2}}{k} . \quad (6.128)$$

This equation is of a form which cannot be solved analytically, therefore we must resort to numerical search methods in order to solve it. In order to see what types of solutions we might get, we can plot the left and right hand sides of this equation as a function of k for a given value of χ . An example is shown in figure 6.4.

You can see that in this example there are discrete solutions, each time that the two curves cross. It is also noticeable that the R.H.S. becomes imaginary for $\chi < k$. This is the point at which the energy of the particle becomes larger than the potential energy of the walls; this will be treated separately below. The lowest energy eigenfunctions for the asymmetric case are shown in figure 6.5

For the *symmetric* solution, we follow a similar method. In this case $u_{II}(x) = A \cos(kx)$ and $u_I(x) = u_{III}(x)$. We therefore know immediately that $a_I = b_{III}$. This results in a simple relation between B , a_I and b_{III}

$$a_I e^{-\kappa a/2} = b_{III} e^{-\kappa a/2} = A \cos(ka/2) . \quad (6.129)$$

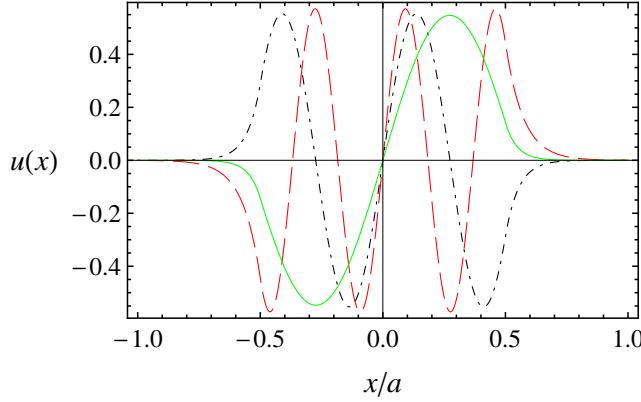


Figure 6.5: Lowest energy asymmetric energy eigenfunctions for the finite square well, for $\chi a = 7\pi$.

The solutions will therefore be

$$u_I(x) = A \cos(ka/2) e^{\kappa(a/2+x)} \quad (6.130)$$

$$u_{II}(x) = A \cos(kx) \quad (6.131)$$

$$u_{III}(x) = A \cos(ka/2) e^{\kappa(a/2-x)} \quad (6.132)$$

and again we have a transcendental equation which must be solved for k

$$\tan(ka/2) = \frac{\sqrt{\chi^2 - k^2}}{k} . \quad (6.133)$$

Again we have no analytical solution. I have plotted the two sides of this equation in figure 6.6. Again you can see a finite number of discrete crossing points, leading to a finite number

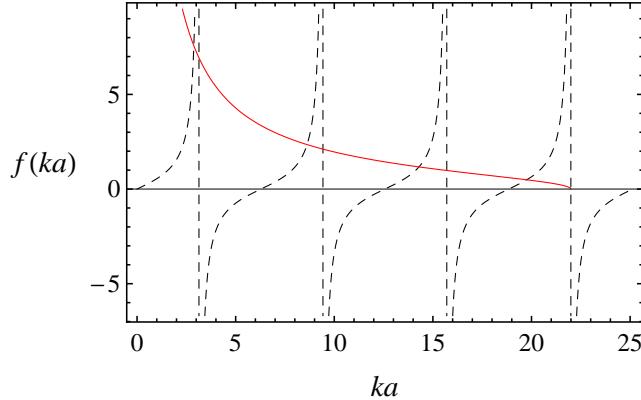


Figure 6.6: L.H.S. (dashed, black) and R.H.S. (solid, red) of equation 6.133 for $\chi a = 7\pi$.

of solutions. For energies where $k > \chi$, we perform a separate analysis below. The first few symmetric eigenstates are plotted in figure 6.7

Comments

Note again the features of these wave functions and energy levels. The energy levels of this problem are discrete due to the boundary conditions (which are only satisfied for solutions of equations 6.128 and 6.133). This is because the particle is confined to a region of space - it is a bound state of the problem.

The ground state is symmetric, and has no zero-crossing points, whereas each successive

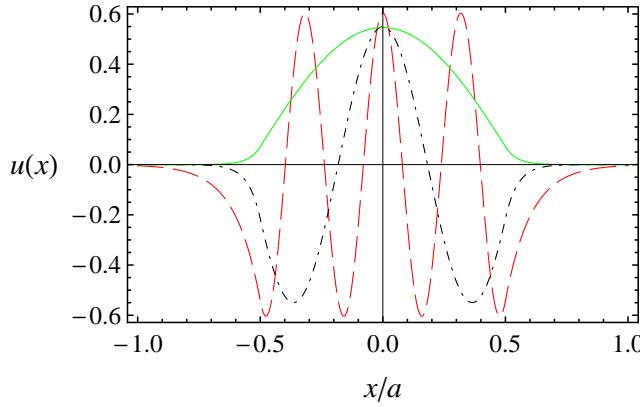


Figure 6.7: Lowest energy symmetric energy eigenfunctions for the finite square well, for $\chi a = 7\pi$.

excited state has one more zero-crossing. The states with higher kinetic energy have higher curvatures, as was found for the infinite square well.

Solutions for $E - V_0 > 0$

If the energy of the particle is greater than the potential of the walls, then for $|x| > a/2$ we have solutions

$$u_I(x) = a_I e^{i\kappa x} + b_I e^{-i\kappa x} \text{ for } x < -a/2 \quad (6.134)$$

$$u_{III}(x) = a_{III} e^{i\kappa x} + b_{III} e^{-i\kappa x} \text{ for } x > a/2 \quad (6.135)$$

with

$$\kappa = \sqrt{\frac{2m}{\hbar^2}(E - V_0)} . \quad (6.136)$$

These must satisfy boundary conditions with the solution inside the well, which is of the form

$$u_{II}(x) = A \cos(kx) + B \sin(kx) \quad (6.137)$$

as before.

First let us consider the boundary conditions for the *anti-symmetric* solution. In this case $u_{II}(x) = B \sin(kx)$ and $u_I(-x) = -u_{III}(x)$. We therefore know immediately that $a_I = -b_{III}$ and $b_I = -a_{III}$. This allows us to treat only one of the boundaries. Choosing the boundary at $x = -a/2$ we have

$$u_I(-a/2) = u_{II}(-a/2) \rightarrow a_I e^{-i\kappa a/2} + b_I e^{i\kappa a/2} = B \sin(k(-a/2)) \quad (6.138)$$

$$\left. \frac{\partial u_I}{\partial x} \right|_{-a/2} = \left. \frac{\partial u_{II}}{\partial x} \right|_{-a/2} \rightarrow i\kappa (a_I e^{-i\kappa a/2} - b_I e^{i\kappa a/2}) = -kB \cos(k(-a/2)) \quad (6.139)$$

which can be solved for a_I and b_I in terms of B . B can then be found by making sure that the wave function is normalized (this is again a case of quasi-normalization, since the plane wave solutions extend throughout the whole of space). The complete treatment yields a lot of algebra but no further physical insight, so we will stop at this point and move on to more important problems.

Comments

It is worth noticing that for this case the wavelength of the solution is different depending on whether the particle is in the region with zero potential, or within the walls. This would be expected due to the reduced kinetic energy within the walls.

Also note that in this case the solutions have a continuous energy spectrum (there is no condition which is equivalent to equations 6.128 or 6.133). Mathematically the discrete spectrum arose in the $E < V_0$ case because we got rid of the exponentially growing solutions in the potential walls, to ensure the wavefunction did not have infinite energy. This meant that we only required one of the boundary conditions to solve for the relative amplitudes of the wavefunction inside and outside the potential well. We then used the other boundary condition to find the condition which k and κ had to satisfy in order to obtain a solution.

For the $E > V_0$ case, we had to solve for an extra constant in getting the solution. This means that there is one extra variable, and thus we had to use both boundary conditions to solve this case. This left no extra constraint on the allowed energies, and hence we have a continuous spectrum.

It is generally true that bound solutions, for which $E < V_0$ at $x = \infty$, have a discrete set of energy eigenvalues, while unbound solutions (for which $E > V_0$ at $x = \infty$) have a continuous spectrum of energy eigenvalues.

6.7 Tunneling

In the previous sections we saw that even if the particle energy is less than the potential in the walls, the wavefunction is non-zero inside the walls. This means that the particle has a probability of being found where it could not possibly appear in classical mechanics! Let us now look at the problem of a potential barrier, as depicted in figure 6.8. The potential barrier has a height V_0 and length L . We will take it to start at $x = 0$. Let us see if we can use solutions of Schrödinger's equation to get a quantitative result.

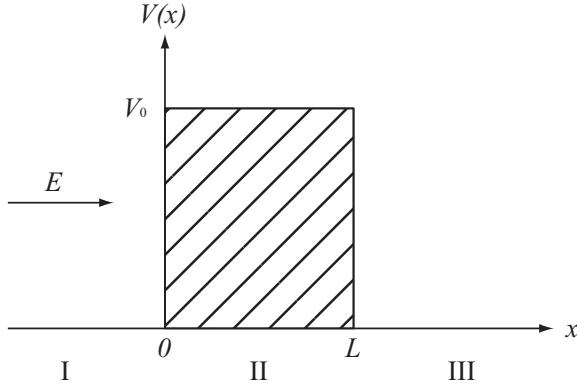


Figure 6.8: A particle approaches a barrier of height V_0 from the left. The particle initially has energy $E < V_0$.

As for the finite square well, we find that the solution to Schrödinger's equation within the region $0 < x < L$ is (for $E - V_0 < 0$)

$$u_{II} = a_{II}e^{\kappa x} + b_{II}e^{-\kappa x} \quad (6.140)$$

where

$$\kappa = \sqrt{\frac{2m}{\hbar^2}(V_0 - E)} . \quad (6.141)$$

In the region where the potential is zero the solutions to the TISE are

$$u_I = a_I e^{ikx} + b_I e^{-ikx} \quad (6.142)$$

$$u_{III} = a_{III} e^{ikx} + b_{III} e^{-ikx} \quad (6.143)$$

where we have now chosen to write these solutions as exponentials rather than cosine or sine. k is given by

$$k = \sqrt{\frac{2m}{\hbar^2}E} . \quad (6.144)$$

This form gives a physical insight. Remember that the e^{ikx} terms are eigenfunctions of the momentum operator with positive momentum $p = \hbar k$, whilst the e^{-ikx} terms are eigenfunctions of the momentum operator with negative momentum. Let us consider that on the left hand side we have an incident wave with unit amplitude ($a_I = 1$). In this case then $|b_I|^2$ will be the reflection probability and $|a_{III}|^2$ gives the transmission probability. Since there is no structure to the potential in region III then we shall assume that $b_{III} = 0$.

If you are worried about some of these assumptions, you are right to be. We recently learnt that a plane wave extends over all space, and here I am telling you that the wave is “incident” from the left hand side etc.. You are correct. What I am describing here is a gross simplification. The true situation is that a wavepacket made up of a sum over plane

waves is incident from the left, and we are interested in the propagation through to the right hand side or the reflection. This situation is much more complex than I'd like to treat here, so I have (over)simplified it to try to give you some insight into the physics without us having to do too much algebra.

We are interested in the transmission probability. This can be defined as

$$T = \frac{|a_{III}|^2}{|a_I|^2}. \quad (6.145)$$

We can find this ratio using the same boundary conditions as we obtained previously, namely that at the boundaries $x = 0$, $x = L$ the wavefunction is continuous in value, and the gradient of the wavefunction is continuous. Mathematically, we find that continuity of the wavefunction gives

$$u_I(0) = u_{II}(0) \rightarrow 1 + b_I = a_{II} + b_{II} \quad (6.146)$$

$$u_{II}(L) = u_{III}(L) \rightarrow a_{III}e^{\kappa L} + b_{III}e^{-\kappa L} = a_{III}e^{ikL} \quad (6.147)$$

and continuity of the gradient of the wavefunction gives

$$\frac{\partial u_I}{\partial x}(0) = \frac{\partial u_{II}}{\partial x}(0) \rightarrow ik - ikb_I = \kappa a_{II} - \kappa b_{II} \quad (6.148)$$

$$\frac{\partial u_{II}}{\partial x}(L) = \frac{\partial u_{III}}{\partial x}(L) \rightarrow \kappa a_{III}e^{\kappa L} - \kappa b_{III}e^{-\kappa L} = ika_{III}e^{ikL} \quad (6.149)$$

From this set of equation coefficients b_I, a_{II}, b_{II} can be eliminated (this is painful) until we obtain an expression for a_{III} , which looks like

$$\frac{1}{a_{III}} = \left(\frac{1}{2} + \frac{\kappa}{4ik} + \frac{ik}{4\kappa}\right)e^{(ik-\kappa)L} + \left(\frac{1}{2} - \frac{\kappa}{4ik} - \frac{ik}{4\kappa}\right)e^{(ik+\kappa)L}. \quad (6.150)$$

You can see that this already looks pretty complicated, despite the drastic simplifications which I made to the problem. Under restricted conditions we can make approximations that will give us a simpler form which still contains the important physical result.

Let us first consider a particle which has a much lower energy than the potential barrier $E \ll V_0$. In this limit $k \ll \kappa$ and we can make the approximation

$$\frac{\kappa}{4k} \pm \frac{k}{4\kappa} \simeq \frac{\kappa}{4k}. \quad (6.151)$$

We can further restrict the problem by assuming that the barrier is wide compared to the de Broglie wavelength of the particle $\kappa L \gg 1$, giving

$$e^{\kappa L} \gg e^{-\kappa L}. \quad (6.152)$$

Under these two approximations, equation 6.150 simplifies to

$$\frac{a_I}{a_{III}} = \left(\frac{1}{2} + \frac{i\kappa}{4k}\right)e^{(ik+\kappa)L}. \quad (6.153)$$

and we obtain the transmission probability

$$T = \frac{|a_{III}|^2}{|a_I|^2} \sim \frac{16}{4 + (\kappa/k)^2} e^{-2\kappa L}. \quad (6.154)$$

If we replace k and κ by their expression in terms of E and V_0 we find

$$T = \frac{16}{3 + V_0/E} e^{-2\sqrt{2m(V_0-E)}L/\hbar}. \quad (6.155)$$

We see that there is a finite transmission probability for the particle to travel through the potential barrier, even though classically it would not have the energy to do this! We call this effect *tunneling*, and it plays a role in a wide range of physics. This is a quantum mechanical effect which is clearly not possible classically.

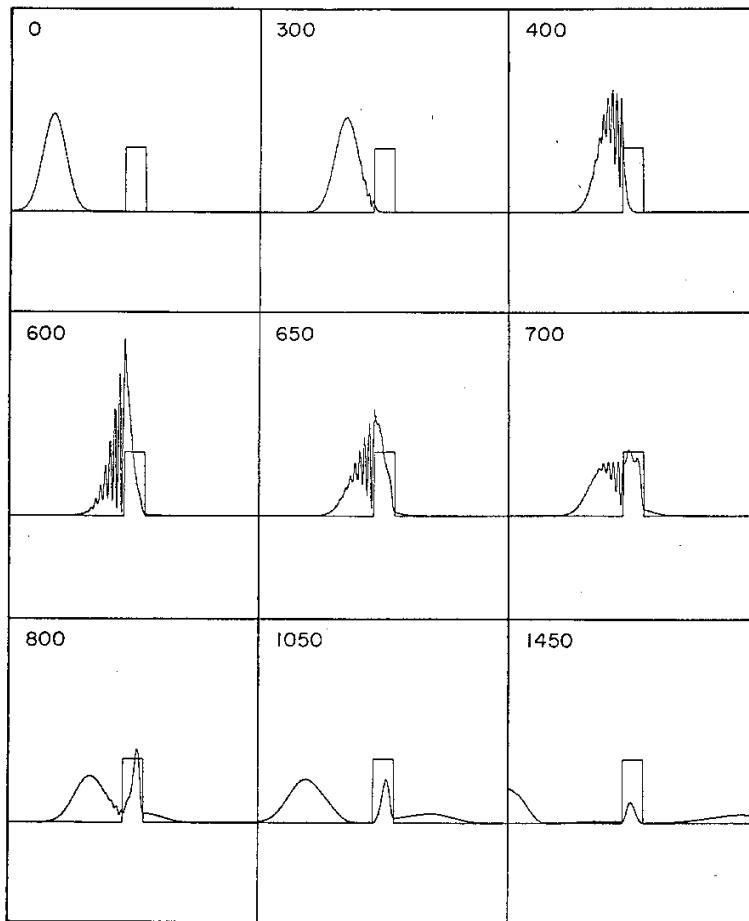


Figure 6.9: A series of plots from a numerical simulation of a gaussian wavepacket scattered from a potential barrier. In this example the mean energy of the wave packet is equal to the potential height. The pictures show how the packet is reflected, that a part is transmitted, and there is also a resonance which decays with time. This picture series is one of the first computer generated numerical solutions to the problem. (A. Goldberg, H. Schey, and J. L. Schwartz, "Computer-Generated Motion Pictures of One-Dimensional Quantum Mechanical Transmission and Reflection Phenomena," Am. J. Phys., 35(3), 1967)

To fill out the full picture of scattering beyond the plane-wave assumptions which we made in this section, figure 6.9 shows the result of numerical simulations of a wave packet scattered by a potential barrier. If you wish, there is an optional numerical problem on the problem set for you to try this out.

Technological applications and limitations

What does this mean for technology? Actually, it turns out to be remarkably useful in some situations, and a bug in others. One example of a device which uses this effect is the “scanning tunneling microscope” (STM), which was invented in 1981 by Gerd Binnig and Heinrich Rohrer at IBM Zürich and earned them the Nobel Prize in 1986. In this device a metallic tip is placed close to the surface of a material which is being studied. The

gap between the tip forms a potential barrier which is higher in energy than the energy of electrons inside the metal. Hence classically they would not have the energy to pass. However electrons are quantum particles, so they may tunnel between the tip and the material. The rate of tunneling is related (by the exponential in equation 6.155) to the distance between the tip and the material, hence the profile of the surface can be mapped out. This method is able to map out surfaces at length scales close to those of the size of individual atoms. An example image from an STM is shown in figure 6.10.

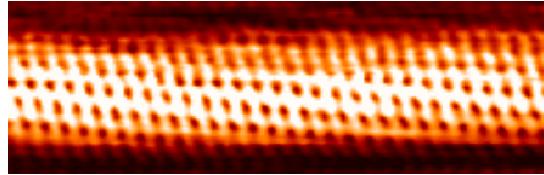


Figure 6.10: An “image” of the surface of a carbon nanotube, taken using a scanning tunneling microscope. The image is a reconstruction made from monitoring the height of the scanning tip above the material surface as a function of time (which is related to the position in the plane of the surface). The individual carbon atoms can clearly be resolved.

Why could tunneling be a bug? In trying to improve computer chips, one approach is to make the physical systems which are used to store the basic units of information smaller. The ultimate limit would be to have a single electron in a potential well representing the binary number “1”, and the absence of an electron representing the binary number “0”. Such devices have been built, with the potential well formed by structuring the material used for the chip. However if the walls of the potential well are made thin, an electron in the potential well has some probability to tunnel out through the sides. This results in a computation error, since our bit flips from “0” to “1”. So tunneling effects could be expected to limit the rapid progress of computers at some point in the future.

6.8 The Quantum Harmonic Oscillator

The quantum harmonic oscillator is one of the most important problems in quantum mechanics. The reason for this is its ubiquity - as in classical mechanics, the leading order terms in a Taylor expansion of a potential about an equilibrium position are quadratic. We can see this from the following expansion

$$V(x_e + \Delta x) = V(x_e) + \frac{\partial V}{\partial x} \Big|_{x_e} \Delta x + \frac{\partial^2 V}{\partial x^2} \Big|_{x_e} (\Delta x)^2 + O(\Delta x^3) \quad (6.156)$$

where x_e is the equilibrium position. In equilibrium, the forces on the object under study are balanced, and thus $\partial V/\partial x|_{x_e} = 0$. The first term in the expansion is just a potential offset which does not affect the dynamics, therefore the leading term which does affect the dynamics is quadratic.

In order to keep things simple, we will treat the 1-D case. We will write the potential term for the oscillator as

$$V(x) = \frac{1}{2}m\omega^2 x^2 \quad (6.157)$$

where m is the mass of the particle in the potential, and ω is a constant (which as you might guess will turn out to be the frequency of the oscillator). Including the usual term for the kinetic energy, the Hamiltonian for the oscillator is

$$\hat{H} = \frac{\hat{p}_x^2}{2m} + \frac{1}{2}m\omega^2 x^2 \quad (6.158)$$

We wish to find the energy eigenfunctions of the oscillator, so are looking to find solutions of the TISE with this Hamiltonian. Writing the momentum operator in terms of position, and taking $\hbar\omega/2$ outside the brackets,

$$\hat{H} = \frac{\hbar\omega}{2} \left(-\frac{\hbar}{m\omega} \frac{\partial^2}{\partial x^2} + \frac{m\omega}{\hbar} x^2 \right) . \quad (6.159)$$

This form encourages us to define a dimensionless parameter

$$\chi \equiv \sqrt{\frac{m\omega}{\hbar}} x \quad (6.160)$$

reducing the Hamiltonian to the simplified form

$$\hat{H} = \frac{\hbar\omega}{2} \left(-\frac{\partial^2}{\partial \chi^2} + \chi^2 \right) . \quad (6.161)$$

The time-independent Schrödinger equation can now be written in the form

$$\frac{\hbar\omega}{2} \left(-\frac{\partial^2}{\partial \chi^2} + \chi^2 \right) \psi(\chi) = E\psi(\chi) . \quad (6.162)$$

Eigenfunctions of the TISE for this problem can be found by a number of routes. One method is to insert a power series

$$\psi(\chi) = \sum_s^\infty a_s \chi^s \quad (6.163)$$

and equate co-efficients of χ^s for different values of s . This is called the Frobenius method, and we will take this type of approach when we solve the radial part of the Schrödinger equation for the Hydrogen atom. However for the case we are considering here there is an alternative and very powerful approach due to Dirac which provides a beautiful framework for dealing with this type of problem. As a result I think it is worth describing here.

6.8.1 Ladder operators

Let us start by multiplying out the following two brackets

$$\left(-\frac{\partial}{\partial\chi} + \chi\right) \left(\frac{\partial}{\partial\chi} + \chi\right) = \left(-\frac{\partial^2}{\partial\chi^2} + \chi^2\right) - \frac{\partial}{\partial\chi}\chi + \chi\frac{\partial}{\partial\chi} \quad (6.164)$$

$$= \left(-\frac{\partial^2}{\partial\chi^2} + \chi^2\right) - \left[\frac{\partial}{\partial\chi}, \chi\right] \quad (6.165)$$

$$= \left(-\frac{\partial^2}{\partial\chi^2} + \chi^2\right) - 1 \quad (6.166)$$

where the square brackets are used to indicate a commutator, and in making the transition between the second and third lines I noted that the commutation relation for χ and $\partial/\partial\chi$ is

$$\left[\frac{\partial}{\partial\chi}, \chi\right] = 1 . \quad (6.167)$$

Now compare equation 6.166 to the form of the Hamiltonian in equation 6.161. The similarity gave Dirac the idea to introduce two new operators defined as

$$\hat{a} = \frac{1}{\sqrt{2}} \left(\frac{\partial}{\partial\chi} + \chi \right) \quad (6.168)$$

$$\hat{a}^\dagger = \frac{1}{\sqrt{2}} \left(-\frac{\partial}{\partial\chi} + \chi \right) \quad (6.169)$$

which we call the annihilation (\hat{a}) and creation (\hat{a}^\dagger) operators for reasons which you will see soon. Using these operators, the Schrödinger equation can be written as

$$\hbar\omega \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) \psi = E\psi . \quad (6.170)$$

At this point we note that since all but one of the terms are constant, the eigenstates of the Hamiltonian will be the same as the eigenstates of the operator $\hat{a}^\dagger \hat{a}$.

It will be useful to evaluate the commutation relation between the operators \hat{a} and \hat{a}^\dagger , which is

$$\begin{aligned} [\hat{a}, \hat{a}^\dagger] &= \left[\frac{1}{\sqrt{2}} \left(\frac{\partial}{\partial\chi} + \chi \right), \frac{1}{\sqrt{2}} \left(-\frac{\partial}{\partial\chi} + \chi \right) \right] \\ &= \frac{1}{2} \left[\frac{\partial}{\partial\chi}, \chi \right] - \frac{1}{2} \left[\chi, \frac{\partial}{\partial\chi} \right] \\ &= \frac{1}{2} (1 - (-1)) = 1 . \end{aligned} \quad (6.171)$$

Note that this allows us to write

$$\hat{a}^\dagger \hat{a} = \hat{a} \hat{a}^\dagger - 1 \quad (6.172)$$

$$\hat{a} \hat{a}^\dagger = \hat{a}^\dagger \hat{a} + 1 . \quad (6.173)$$

This may look trivial to some of you, but these relations will be useful to us later on.

6.8.2 Eigenstates

How can these operators help us to find the energy eigenstates? First we note that the eigenstates of $\hat{a}^\dagger \hat{a}$ are the same as the eigenstates of the Hamiltonian. The eigenstate ψ_n of the operator $\hat{a}^\dagger \hat{a}$ satisfies the equation

$$\hat{a}^\dagger \hat{a} \psi_n = n\psi_n \quad (6.174)$$

where n is the eigenvalue. Let us now apply the operator \hat{a} to both sides of this equation. We find

$$\hat{a}(\hat{a}^\dagger \hat{a})\psi_n = \hat{a}n\psi_n . \quad (6.175)$$

Now we should notice that the operator $\hat{a}\hat{a}^\dagger \hat{a}$ which is found on the left hand side of this equation can be transformed using equation 6.173 into

$$\hat{a}\hat{a}^\dagger \hat{a} = (\hat{a}^\dagger \hat{a} + 1)\hat{a} . \quad (6.176)$$

This allows us to transform equation 6.175 to

$$(\hat{a}^\dagger \hat{a} + 1)\hat{a}\psi_n = n\hat{a}\psi_n . \quad (6.177)$$

Re-arranging this expression allows us to write a new eigenvalue equation for the operator $\hat{a}^\dagger \hat{a}$

$$\hat{a}^\dagger \hat{a}(\hat{a}\psi_n) = (n - 1)(\hat{a}\psi_n) . \quad (6.178)$$

with eigenvalue $n - 1$ and eigenstate $\hat{a}\psi_n$. This eigenstate satisfies the same equation but with n reduced by exactly 1. We can thus define it as the eigenstate

$$\psi_{n-1} = N_n \hat{a}\psi_n \quad (6.179)$$

where N_n is a constant which ensures that the state is still normalized.

Since the eigenstates of $\hat{a}^\dagger \hat{a}$ are the same as the energy eigenstates, this equation tells us that if we can find one energy eigenstate, we can find a whole chain of energy eigenstates with lower energy by repeatedly applying \hat{a} .

Let us also consider a similar approach but now applying \hat{a}^\dagger to equation 6.174. We find that

$$\begin{aligned} \hat{a}^\dagger(\hat{a}^\dagger \hat{a})\psi_n &= \hat{a}^\dagger n\psi_n \\ \hat{a}^\dagger(\hat{a}\hat{a}^\dagger - 1)\psi_n &= n\hat{a}^\dagger \psi_n \end{aligned} \quad (6.180)$$

which on re-arranging gives us

$$\hat{a}^\dagger \hat{a}(\hat{a}^\dagger \psi_n) = (n + 1)\hat{a}^\dagger \psi_n . \quad (6.181)$$

Again this is an eigenvalue equation, but this time with eigenvalue $n + 1$ and with eigenstate

$$\psi_{n+1} = K_n \hat{a}^\dagger \psi_n \quad (6.182)$$

where K_n is again a constant inserted to ensure that the new state is normalized.

6.8.3 The ground state of energy

We now have a systematic way of generating new eigenstates of $\hat{a}^\dagger \hat{a}$ so long as we are given one of them. As a starting point, we need to find one of the states. For the harmonic oscillator, we note that the energy can only be positive (kinetic energy can only be positive, and the potential is also also positive because it is quadratic). The ground state will be the lowest of these in energy, therefore there can be no state with a lower value of n . However we know that we can always generate an eigenstate with a lower value of n using equation 6.179. In order to have a lowest energy state we thus require that for the minimum value of n the eigenstate satisfies

$$\hat{a}\psi_0 = 0 . \quad (6.183)$$

Such a state satisfies the eigenvalue equation with $n = 0$

$$\hat{a}^\dagger \hat{a} \psi_0 = 0\psi_0 \quad (6.184)$$

which means that the energy of the state is

$$E_0 = \frac{\hbar\omega}{2} . \quad (6.185)$$

The lowest energy state of the harmonic oscillator still has energy! This is called the zero-point energy. For any system which can be accurately described by a harmonic oscillator there is always energy present!

Now let us find the form of the lowest energy eigenstate. We can do this by writing 6.183 in terms of χ , giving a simple differential equation

$$\frac{\partial \psi_0(\chi)}{\partial \chi} = -\chi \psi_0 . \quad (6.186)$$

To find the solution to this equation we take

$$\frac{d\psi_0}{\psi_0} = -\chi d\chi \quad (6.187)$$

which on integrating both sides and taking the exponential, gives

$$\psi_0(\chi) = N e^{-\chi^2/2} \quad (6.188)$$

where N is a constant which can be determined by normalization. The normalization can be performed using one of the standard integrals

$$\begin{aligned} \int_{-\infty}^{\infty} \psi_0^*(\chi) \psi_0(\chi) d\chi &= |N|^2 \int_{-\infty}^{\infty} e^{-\chi^2} d\chi \\ &= |N|^2 \sqrt{\pi} \end{aligned} \quad (6.189)$$

giving

$$N = \pi^{-1/4} . \quad (6.190)$$

The ground state of the quantum harmonic oscillator is thus

$$\psi_0(\chi) = \pi^{-1/4} e^{-\chi^2/2} \quad (6.191)$$

which can be written in terms of x as

$$\psi_0(x) = \left(\frac{m\omega}{\pi\hbar} \right)^{1/4} e^{-m\omega x^2/(2\hbar)} \quad (6.192)$$

6.8.4 Energy levels

Written in terms of the creation and annihilation operators, the Schrödinger equation for the Harmonic oscillator is of the form

$$\hbar\omega \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) \psi = E\psi . \quad (6.193)$$

The eigenfunctions of $\hat{a}^\dagger \hat{a}$ are therefore the same as the eigenstates of the Hamiltonian, and we can write the energy levels as

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right) \quad (6.194)$$

As we found above, the ground state ($n = 0$) has finite zero point energy of $\hbar\omega/2$. This is certainly different from the classical case - a stationary classical oscillator in its equilibrium position does have zero energy. The higher energy states are at energies which are integer multiples of $\hbar\omega$ above the ground state, the first few being $3/2\hbar\omega, 5/2\hbar\omega, 7/2\hbar\omega$ etc..

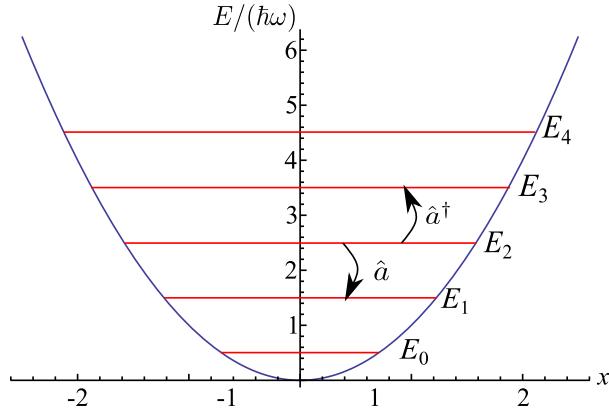


Figure 6.11: The energy levels of the harmonic oscillator are equally spaced in energy by $\hbar\omega$. The ladder operators enable us to find the eigenstate of the harmonic oscillator Hamiltonian if we have one of them already. In this sense they “create” and “annihilate” quanta of energy.

6.8.5 The excited states

Now let us use \hat{a}^\dagger and \hat{a} to find the form of the first couple of excited states. Applying \hat{a}^\dagger to ψ_0 , we find that

$$\begin{aligned}\hat{a}^\dagger\psi_0 &= \frac{1}{\sqrt{2}} \left(-\frac{\partial}{\partial \chi} + \chi \right) \pi^{-1/4} e^{-\chi^2/2} \\ &= \frac{1}{\sqrt{2\sqrt{\pi}}} (\chi e^{-\chi^2/2} + \chi e^{-\chi^2/2}) \\ &= \sqrt{\frac{2}{\sqrt{\pi}}} \chi e^{-\chi^2/2}\end{aligned}\tag{6.195}$$

This is not exactly the first excited state, because we haven't shown whether the \hat{a}^\dagger operator preserves normalization (we only showed that it produced a new eigenstate of $\hat{a}^\dagger \hat{a}$, not a normalized eigenstate). In general it does not, but in this case

$$\int_{-\infty}^{\infty} \frac{2}{\sqrt{\pi}} \chi^2 e^{-\chi^2} d\chi = 1\tag{6.196}$$

and therefore

$$\psi_1(\chi) = \sqrt{\frac{2}{\sqrt{\pi}}} \chi e^{-\chi^2/2}\tag{6.197}$$

For the second excited state, we can follow the same procedure, and obtain

$$\hat{a}^\dagger \psi_1(\chi) = \sqrt{\frac{1}{\sqrt{\pi}}} (2\chi^2 - 1) e^{-\chi^2/2}\tag{6.198}$$

Let us again check the normalization. We find that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} (2\chi^2 - 1)^2 e^{-\chi^2} d\chi = 2\tag{6.199}$$

implying that the properly normalized eigenstate is

$$\psi_2(\chi) = \frac{1}{\sqrt{2}} \hat{a}^\dagger \psi_1(\chi) = \sqrt{\frac{1}{2\sqrt{\pi}}} (2\chi^2 - 1) e^{-\chi^2/2}\tag{6.200}$$

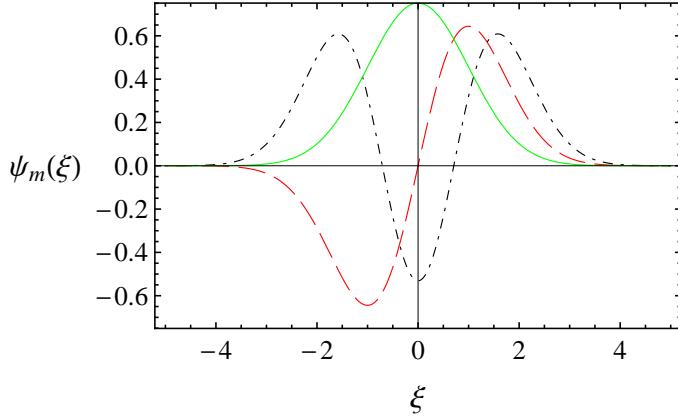


Figure 6.12: Lowest energy eigenstates of the harmonic oscillator, plotted in terms of the scaled space parameter χ . The ground state: green, solid, first excited state: red, dashed, and second excited state: black, dot-dashed.

The general rule for the normalization of these states is that

$$\hat{a}^\dagger \psi_n(\chi) = \sqrt{n+1} \psi_{n+1}(\chi) \quad (6.201)$$

$$\hat{a} \psi_n(\chi) = \sqrt{n} \psi_{n-1}(\chi) . \quad (6.202)$$

By successive application of equation 6.201 we can find that the general form for the excited state with eigenvalue of \hat{a}^\dagger equal to n is given by

$$\psi_n(\chi) = \left(\frac{1}{\pi 2^{2n} (n!)^2} \right)^{1/4} e^{-\chi^2/2} H_n(\chi) \quad (6.203)$$

where $H_n(\chi)$ is a polynomial in χ called a *Hermite* polynomial. The Hermite polynomials for the first few excited states are

$$\begin{aligned} H_0(\chi) &= 1 \\ H_1(\chi) &= 2\chi \\ H_2(\chi) &= -2(1 - 2\chi^2) \\ H_3(\chi) &= -12 \left(\chi - \frac{2}{3}\chi^3 \right) \end{aligned} \quad (6.204)$$

Note that the exponential factor is positive for all x . It is the form of the Hermite polynomials which therefore determines the number of times the wavefunction crosses zero. On inspection of these polynomials, we see the usual rule, which is that H_0 has no zero crossings, H_1 has one, and so on.

To illustrate this, the lowest three energy eigenstates $n = 0, 1, 2$ of the harmonic oscillator are shown in figure 6.12 written as a function of χ . Excited states with $n = 50$ and $n = 60$ are shown in figure 6.13. Note here that the oscillator has a higher probability of being found at the extremes of its motion - this is a feature that you are familiar with from classical mechanics (where the oscillator spends more time at the extremes of oscillation because it has lower momentum there). Nevertheless, it is important to remind ourselves that the energy eigenstates of the harmonic oscillator are stationary states. They evolve in time according to a physically insignificant phase factor and the probability to find the oscillator at a fixed position *does not change with time*. This is very different from the classical oscillator, which has a fixed energy but oscillates back and forth.

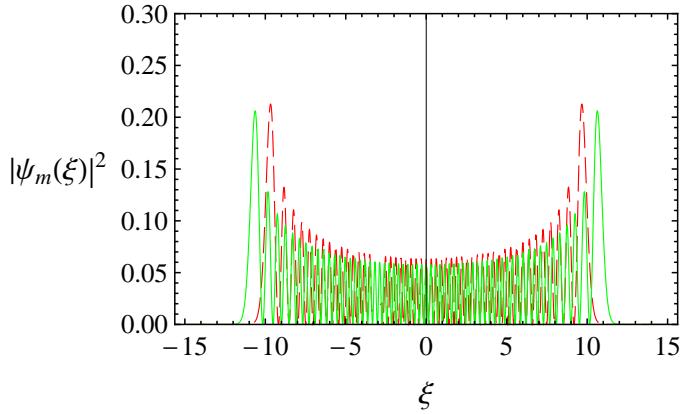


Figure 6.13: The probability distributions for the $n = 50$: red, dashed and $n = 60$: solid, green, energy eigenfunctions of the harmonic oscillator. You should note that the probability of finding the particle at high amplitudes of oscillation is higher than the probability of finding it in the center.

6.8.6 Expectation values of the harmonic oscillator

The creation and annihilation operators are extremely powerful, and are used a lot in modern quantum mechanics. This is in part because of the fact that the harmonic oscillator problem appears in a wide range of physical situations. We use the harmonic oscillator to describe photons of light, phonons (vibrations in solids) and similar techniques are used for electrons in matter. As an example of the power of using ladder operators for these types of problems, I will show how to use them to evaluate expectation values of the physical variables for position and momentum for all eigenstates of energy.

In this context it is useful to relate the operators for \hat{x} and \hat{p} to the ladder operators \hat{a}^\dagger and \hat{a} . We can do this by writing the latter in terms of x using equations 6.168 and 6.169, giving

$$\begin{aligned}\hat{a}^\dagger &= -\frac{1}{\sqrt{2}} \left(-\sqrt{\frac{\hbar}{m\omega}} \frac{\partial}{\partial x} + \sqrt{\frac{m\omega}{\hbar}} x \right) \\ &= -i \left(\frac{1}{2m\omega\hbar} \right)^{1/2} \hat{p} + \left(\frac{m\omega}{2\hbar} \right)^{1/2} \hat{x} .\end{aligned}\quad (6.205)$$

$$\begin{aligned}\hat{a} &= -\frac{1}{\sqrt{2}} \left(\sqrt{\frac{\hbar}{m\omega}} \frac{\partial}{\partial x} + \sqrt{\frac{m\omega}{\hbar}} x \right) \\ &= i \left(\frac{1}{2m\omega\hbar} \right)^{1/2} \hat{p} + \left(\frac{m\omega}{2\hbar} \right)^{1/2} \hat{x} .\end{aligned}\quad (6.206)$$

By combining these two expressions we are able to obtain definitions of \hat{x} and \hat{p} in terms of the ladder operators

$$\hat{x} = \left(\frac{\hbar}{2m\omega} \right)^{1/2} (\hat{a}^\dagger + \hat{a}) \quad (6.207)$$

$$\hat{p} = i \left(\frac{m\hbar\omega}{2} \right)^{1/2} (\hat{a}^\dagger - \hat{a}) . \quad (6.208)$$

These definitions are useful because they allow us to evaluate properties such as expectation values for harmonic oscillator states without performing any integral. As simple examples we will use them to calculate $\langle \hat{x} \rangle$, $\langle \hat{p} \rangle$, $\langle \hat{x}^2 \rangle$, $\langle \hat{p}^2 \rangle$.

Let us start by evaluating $\langle \hat{x} \rangle$ for the energy eigenstate ψ_n . This is done by evaluating the

integral

$$\int_{-\infty}^{\infty} \psi_n^* x \psi_n dx = \left(\frac{\hbar}{2m\omega} \right)^{1/2} \int_{-\infty}^{\infty} \psi_n^* (\hat{a}^\dagger + \hat{a}) \psi_n dx \quad (6.209)$$

$$= \left(\frac{\hbar}{2m\omega} \right)^{1/2} \left[\int_{-\infty}^{\infty} \psi_n^* \hat{a}^\dagger \psi_n dx + \int_{-\infty}^{\infty} \psi_n^* \hat{a} \psi_n dx \right] \quad (6.210)$$

In order to evaluate the two integrals, we first remember the action of the \hat{a}^\dagger and \hat{a} operators on the eigenstates, which are given by equations 6.202 and 6.201. We find

$$\langle \hat{x} \rangle = \left(\frac{\hbar}{2m\omega} \right)^{1/2} \left[\sqrt{n+1} \int_{-\infty}^{\infty} \psi_n^* \psi_{n+1} dx + \sqrt{n} \int_{-\infty}^{\infty} \psi_n^* \psi_{n-1} dx \right] . \quad (6.211)$$

Since the eigenfunctions of energy are mutually orthogonal, the integrals both equal zero and we find $\langle \hat{x} \rangle = 0$. The expectation value of the position of the oscillator does not change in time for any eigenstate of the energy operator. We can also notice that \hat{p} is identical to \hat{x} apart from a minus sign, and thus we also know that $\langle \hat{p} \rangle = 0$ too.

Next let us try to evaluate $\langle \hat{x}^2 \rangle$. The operator squared gives

$$\hat{x}^2 = \frac{\hbar}{2m\omega} (\hat{a}^\dagger \hat{a}^\dagger + \hat{a} \hat{a} + \hat{a}^\dagger \hat{a} + \hat{a} \hat{a}^\dagger) \quad (6.212)$$

Since we already looked at a simple example, we have some expectation of what the behavior will be. If we have an operator which changes an eigenstate into one of the other eigenstates, we will get zero. Thus the first two terms in the above expression will give zero. For the second two, we simplify things using the commutation relation in equation 6.171 to write

$$\hat{a}^\dagger \hat{a} + \hat{a} \hat{a}^\dagger = 2 \hat{a}^\dagger \hat{a} + 1 \quad (6.213)$$

and since we already know that $\hat{a}^\dagger \hat{a} \psi_n = n \psi_n$ we can easily evaluate $\langle \hat{x}^2 \rangle$ as

$$\langle \hat{x}^2 \rangle = \int_{-\infty}^{\infty} \psi_n^* x^2 \psi_n dx = \frac{\hbar}{2m\omega} (2n+1) . \quad (6.214)$$

Even in the ground state, the oscillator position is uncertain (the wavefunction takes up a finite region of space). It follows that the zero point energy is not due to the mean position and momentum of the oscillator moving away from zero, but the uncertainty in the position. These are called zero-point fluctuations. It turns out that light in space is well described by a harmonic oscillator. In a vacuum, you might expect that there is no light, but actually there are always zero-point fluctuations. The vacuum has energy!

In a similar fashion, we can find that

$$\langle \hat{p}^2 \rangle = \left(\frac{m\hbar\omega}{2} \right) (2n+1) . \quad (6.215)$$

It is worth pointing out that product of the uncertainties in \hat{x} and \hat{p} is

$$(\langle \hat{x}^2 \rangle - \langle \hat{x} \rangle^2) (\langle \hat{p}^2 \rangle - \langle \hat{p} \rangle^2) = \frac{\hbar^2}{4} (2n+1)^2 \quad (6.216)$$

which for the ground state satisfies exactly the Heisenberg uncertainty principle

$$(\Delta x)^2 (\Delta p)^2 = \frac{\hbar^2}{4} \quad (6.217)$$

6.8.7 Classical/Quantum correspondence

How can we relate the quantum and classical views of the harmonic oscillator? In classical physics an oscillator oscillates, but we just saw that in quantum mechanics the eigenfunctions of energy have zero mean position and momentum, even at high energies. It is clear that if quantum mechanics is more fundamental than classical physics, then we should recover classical physics from a quantum mechanical description. It turns out that this can happen, but the states which we see classically are not necessarily energy eigenstates.

As an example there is a quantum mechanical state which does act like the equivalent classical state (ie. its expectation values for position and momentum oscillate back and forth). We can write it in terms of the eigenstates of energy ψ_n as

$$\phi_\alpha = e^{-|\alpha|^2/2} \sum_n^\infty \frac{\alpha^n}{\sqrt{n!}} \psi_n \quad (6.218)$$

where α is a complex number which is related to the amplitude of oscillation. You will show that the expectation value for position for this state goes as

$$\langle \hat{x} \rangle = \sqrt{\frac{\hbar}{2m\omega}} (\alpha e^{i\omega t} + \alpha^* e^{-i\omega t}) \quad (6.219)$$

which oscillates back and forth with angular frequency ω just as the classical harmonic oscillator does. The expectation value $\langle \hat{x}^2 \rangle$ does not change in time - this is also what we would expect classically (where our particle is not expected to expand or contract).

The states which are written in equation 6.218 are called the *coherent or Glauber states* and they are very famous in the theory of the quantum mechanics of light. Glauber won the Nobel prize for his work on the quantum states of light fairly recently, in 2005. The reason that this work is so important is that light from a laser is very well described by a coherent state.

Chapter 7

Hydrogen

The spectrum of hydrogen was one of the main clues which led to the formation of quantum mechanics. The mystery was, how to explain a discrete set of lines in the spectrum. The advantage of looking at the hydrogen spectrum was its relative simplicity.

Hydrogen is the simplest atom which we can look at. It consists of a proton in the nucleus and a bound electron outside. Because it is so simple, it is a primary testing ground for comparing theoretical calculations and experiments. Even today work continues on precise measurements of the hydrogen spectral lines. The frequency of one of these transitions is known to an accuracy of 14 decimal places ($2\ 466\ 061\ 413\ 187\ 074 \pm 74$ Hz), is one of the primary constraints on fundamental constants of physics such as the fine structure constant (which I'll introduce you to later in the lecture, and you will certainly encounter again).

7.1 Pre-quantum models

7.1.1 Early Hydrogen spectra



Figure 7.1: Visible spectral lines of Hydrogen. These are called the Balmer series, after the Swiss schoolteacher who first came up with a formula for the transition frequencies.

Figure 7.1 shows the visible part of the hydrogen spectrum obtained using a grating spectrograph. And in figure 7.2, the full spectrum is sketched. It can be seen that several series occur, which start with well spaced spectral lines, which get closer and closer towards higher frequencies. A formula for obtaining the lines of the visible spectrum was first produced purely empirically by a Swiss schoolteacher, Johann Balmer. His equation for the positions of the spectral lines is

$$\lambda = \frac{n^2}{n^2 - 4} G \quad (7.1)$$

where G is a constant and the integers $n > 2$ are different for each spectral line. Later work

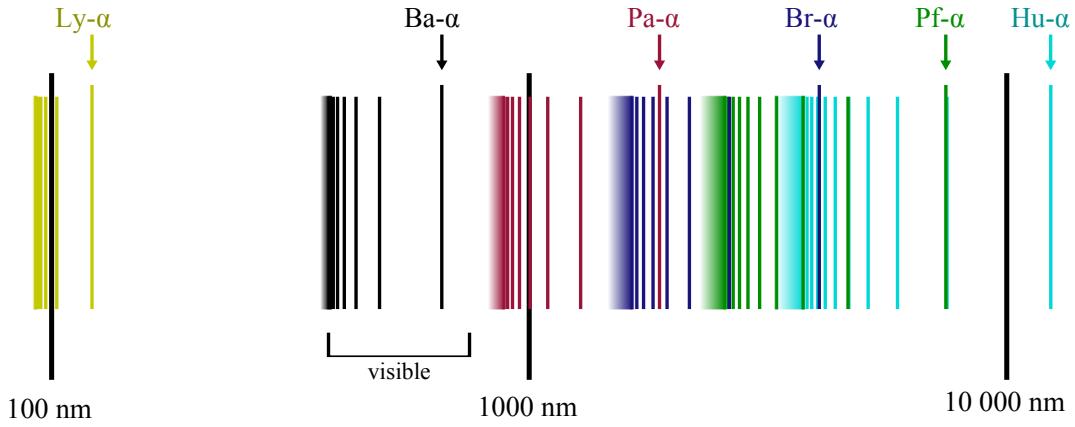


Figure 7.2: Full Hydrogen spectrum on a logarithmic scale. Note the different series, named after Lyman, Balmer, Paschen Graphic from Wikipedia.

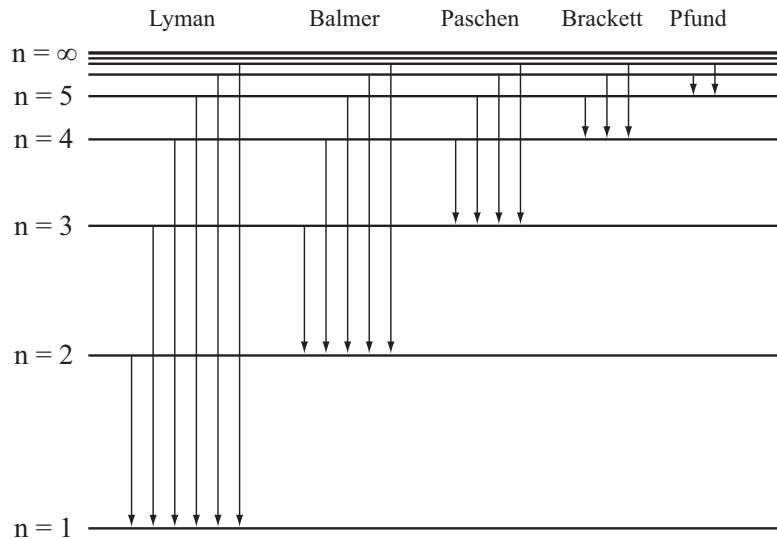


Figure 7.3: Energy diagram of the Hydrogen atom. Included are different sets of transitions, which were discovered independently. The names of these sets of lines are also labeled in figure 7.2. Note that the energy levels get closer together as the energy increases.

by Rydberg led to all of the series being explained by the form

$$\nu = \frac{c}{\lambda} = cR_{\infty} \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) \quad (7.2)$$

where now $n > n'$ and $R_{\infty} = 10973731.568527 \pm 73 \text{ m}^{-1}$ is the Rydberg constant (note the discussion in Hakan and Wolf about the difference between this value and the actual frequencies of the hydrogen transition. I copied the quoted value from the webpage of Prof. Hänsch at MPQ in Munich, who I think has made the most accurate measurements). I have rarely used wavenumbers ($1/\lambda$) since I stopped doing atomic physics at undergraduate level, so this is not a number which is familiar to me. My preference, and a number I can remember off the top of my head, is $R_e = hcR_{\infty} = 13.6 \text{ eV}$ (quoted to much lower accuracy). It was found that the form given in equation 7.2 fits the experiments extremely well (to better than 1 part in 10^5 - see Haken and Wolf for more details).

7.1.2 The Bohr model

So at the beginning of the 1900s the major challenge was to find a theory that could reproduce the data - discrete spectral lines separated according to equation 7.2. The first breakthrough in this regard was from Bohr. I now present Bohr's postulates, with the caveat that only one of them is in use today. Quantum mechanics has now completely superseded the Bohr theory. Nevertheless, they deserve presentation for two reasons. First you must appreciate how big a step into the unknown Bohr took, because that is something which physics need from time to time. Also, Bohr theory provides us with some useful relations which make it easier to remember how parameters scale.

The Bohr postulates were

- The only classical orbits for an electron around a nucleus are *discrete* orbits with energies E_n .
- If an electron is in one of these orbits, it does not emit radiation. An electron can transfer from one orbit to the other, and it will emit light in doing so. The light emitted has frequency $\nu = E_n - E_{n'}$. By comparing this form with equation 7.2 we find that $E_n = -R_e/n^2$ and $E_{n'} = -R_e/n'^2$. Note that a classically orbiting electron is always accelerating, and would thus continuously radiate.
- With increasing orbital radius r the laws of the Bohr model must become identical with classical physics. This is the *correspondence* principle which *is* still in use today. We have tested classical physics very well at macroscopic scales, so it makes sense that quantum theory should reproduce classical behavior in this limit. For Bohr's atom, this means that light should be emitted at the orbital frequency of the electron.

By combining these postulates with the Rydberg formula we are able to derive a value for the Rydberg constant from theory. This was a major success of the day. The working goes as follows.

Classical electron orbit

The classical potential force on the electron at distance r from the nucleus is

$$\frac{e^2}{4\pi\epsilon_0 r^2} \quad (7.3)$$

In order to have stable orbit, this must equal the centrifugal force, which is given by mv^2/r . For these to equal we have

$$v^2 = \frac{e^2}{4\pi\epsilon_0 mr} \quad (7.4)$$

or written in terms of the angular frequency $\omega = v/r$

$$r = \left(\frac{e^2}{4\pi\epsilon_0 m\omega^2} \right)^{1/3}. \quad (7.5)$$

The total energy will be a sum of the potential and kinetic energies

$$U = \frac{mv^2}{2} - \frac{e^2}{4\pi\epsilon_0 r}, \quad (7.6)$$

which on inserting our expression for r and v , we find

$$U = -\frac{1}{2} \frac{e^2}{4\pi\epsilon_0 r} = -\frac{1}{2} \left(\frac{e^4 m}{16\pi^2 \epsilon_0^2} \right)^{1/3} \omega^{2/3}. \quad (7.7)$$

Bohr result - classical limit

The third Bohr postulate states that the Bohr theory should reproduce classical results for large orbits. What predictions should match? Well one would expect from classical theory that an orbiting electron would emit radiation at a frequency corresponding to the orbital frequency of its motion. Therefore Bohr required that transitions between adjacent highly excited levels should also lead to radiation at this same frequency. Using this equivalence, he was able to make a prediction about the value of the Rydberg constant. Let us see how this works.

Taking two large adjacent orbits, for which $n, n' \gg (n - n') = \delta n = 1$, we find

$$\hbar\omega = R_e \left(\frac{1}{n'^2} - \frac{1}{n^2} \right) = R_e \frac{n^2 - n'^2}{n'^2 n^2} \simeq R_e \frac{2n\delta n}{n^4} \quad (7.8)$$

where I have made the approximation that $n \simeq n'$. If the electron jumps between these orbits it will emit a photon with this energy.

The energy of the atom in one of these levels is $U = -R_e/n^2$, therefore we find using equation 7.7 that

$$\frac{R_e}{n^2} = \frac{1}{2} \left(\frac{e^4 m}{16\pi^2 \epsilon_0^2} \right)^{1/3} \left(\frac{2R_e \delta n}{\hbar n^3} \right)^{2/3} \quad (7.9)$$

which after solving for R_e and inputting values for \hbar , the mass of the electron m and the permeability of free space ϵ_0 gives

$$R_e = \left(\frac{e^4 m}{32\pi^2 \epsilon_0^2 \hbar^2} \right) = 13.6057 \text{ eV} \quad (7.10)$$

The ability to make a theoretical prediction of the value of the Rydberg constant which agreed with the measured value was the major success of Bohr's model of the Hydrogen atom. The idea to use the correspondence between microscopic and macroscopic theories is quite remarkable, for me at least, and still gets used in derivations to this day.

What is wrong about the Bohr model?

Though the Bohr model reproduces the energy levels of Hydrogen and was able to predict a value for the Rydberg constant, it doesn't provide any reason for why an electron at a given energy orbiting a nucleus does not emit light. By all the classical theories we know, an accelerated charge should produce radiation (indeed we know that this result is true in quantum theory too). In addition, the Bohr theory makes one prediction which turns out to be wrong. In the Bohr theory, the angular momentum of the electron in the n th energy level is

$$|l| = mv_n r_n = n\hbar \quad (7.11)$$

which for $n = 1$ of course gives \hbar . It has now been shown that in the lowest energy level of the hydrogen atom the electron has an angular momentum of zero - this comes out naturally in the quantum treatment, as we will show below.

The Bohr radius

One useful quantity which we can obtain from the Bohr model is the Bohr radius, which we can obtain by equating the centrifugal force and the Coulomb force

$$\frac{l^2}{2mr^3} = \frac{1}{2} \frac{e^2}{4\pi\epsilon_0 r^2} \quad . \quad (7.12)$$

The smallest orbit will be that with the lowest angular momentum, and we call the radius of this orbit the *Bohr radius*

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} . \quad (7.13)$$

We will see the Bohr radius arising in the quantum treatment.

The fine-structure constant

One of the most useful constants in quantum mechanics can be derived quickly from the Bohr model. For an electron in its lowest energy state ($n = 1$), we look at the value of

$$\alpha = \frac{v}{c} = \frac{\hbar}{ma_0c} = \frac{e^2}{4\pi\epsilon_0\hbar c} . \quad (7.14)$$

This is a dimensionless constant containing all the basic constants. It is just a number, but it combines e from electromagnetism, \hbar from quantum mechanics, and c from relativity. We call this the *fine structure constant*, and it finds use throughout electrodynamics. Its value is roughly 1/137.

One area of research today is to measure whether α is changing. If it was, this would imply that the relationships between the laws of physics aren't constant in time. That the laws of nature are constant with time is a fundamental assumption of our current physical theories, so people perform tests to check that this is the case. A variety of tests have been made, either by making high precision measurements of atomic spectral lines on earth, or by looking at spectral lines of atoms in deep space. So far, I think that every measurement is consistent with α being a constant over time and space, so for the moment, the ground beneath our feet is still solid and we can be reassured that it has been that way throughout the life of the universe. Wikipedia gives a nice review of experimental work in this area.

7.2 Quantum mechanical treatment

The Time-Independent Schrödinger equation for Hydrogen written in the rest frame of the proton and in the approximation that $m \ll m_p$ is

$$\frac{-\hbar^2}{2m} \nabla^2 \psi - \frac{e^2}{4\pi\epsilon_0 r} \psi = E\psi \quad (7.15)$$

where r is the distance between the proton and the electron. We are interested in finding the energy eigenfunctions which satisfy this equation. Since the potential is spherically symmetric, we choose to work in spherical co-ordinates r, θ, ϕ for which

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \quad (7.16)$$

The kinetic energy term of the Schrödinger equation therefore looks like

$$-\frac{\hbar^2}{2mr^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{2mr^2} \left[-\frac{\hbar^2}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) - \frac{\hbar^2}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right]. \quad (7.17)$$

It is worth recalling your classical mechanics at this point. There, when you considered the kinetic energy due to angular momentum, you found it was of the form

$$\frac{l^2}{2I} \quad (7.18)$$

where I is the moment of inertia of the object and l is its angular momentum. For a single particle of mass m a distance r from the point about which it is rotating, the moment of inertia is mr^2 . It therefore makes sense that the term inside the square brackets in equation 7.17 is equivalent to the angular momentum squared, and indeed it is the operator for the angular momentum squared \hat{l}^2 , which we will see in the next section. You can probably already see that this operator acts exclusively on the angular co-ordinates θ and ϕ .

7.2.1 Separation of variables

Since the contribution of the potential energy to the Schrödinger equation is purely radial, we are motivated to search for solutions of the form

$$\psi(r, \theta, \phi) = R(r)Y(\theta, \phi) \quad (7.19)$$

and to see if it possible to use the method of *separation of variables* again to find the solution. Inserting this solution into the Schrödinger equation along with the operator for ∇^2 and the operator \hat{L}

$$\frac{-\hbar^2 Y(\theta, \phi)}{2mr^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} R(r) \right) + \frac{R(r)}{2mr^2} \hat{L}^2 Y(\theta, \phi) = \left(\frac{e^2}{4\pi\epsilon_0 r} + E \right) R(r)Y(\theta, \phi) \quad (7.20)$$

which when we divide through by RY and multiply through by r^2 gives

$$\frac{-\hbar^2}{2mR(r)} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} R(r) \right) + \frac{1}{2mY(\theta, \phi)} \hat{L}^2(Y(\theta, \phi)) = \left(\frac{e^2 r}{4\pi\epsilon_0} + Er^2 \right) \quad (7.21)$$

Since the terms which are dependent on r must be satisfied for all θ, ϕ , and vice versa we can now solve separately for $R(r)$ and $Y(\theta, \phi)$. This we do in the following sections.

7.2.2 Angular momentum eigenfunctions

In order to find angular solutions for the hydrogen atom, we must learn a bit about angular momentum and the way in which we represent it in quantum mechanics. In doing so, we will find the eigenvectors and eigenvalues associated with angular momentum in a problem with spherical symmetry, which we will then apply to our hydrogen problem.

First, let me remind you that angular momentum is given by the vector

$$\mathbf{l} = \mathbf{r} \times \mathbf{p} \quad (7.22)$$

where \mathbf{r} and \mathbf{p} are the position and momentum of the electron. In order to find the operators for angular momentum, we must therefore find three different components, which since they are quantum mechanical operators we shall put hats on. These three components are

$$\hat{l}_x = y\hat{p}_z - z\hat{p}_y \quad (7.23)$$

$$\hat{l}_y = z\hat{p}_x - x\hat{p}_z \quad (7.24)$$

$$\hat{l}_z = x\hat{p}_y - y\hat{p}_x . \quad (7.25)$$

Since we are working with a spherically symmetric problem, let us move to a spherical coordinate system, which is defined as $r = \sqrt{x^2 + y^2 + z^2}$, $\theta = \cos^{-1}(z/r)$ and $\phi = \tan^{-1}(y/x)$, or in the inverse form $x = r \sin \theta \cos \phi$, $y = r \sin \theta \sin \phi$ and $z = r \cos \theta$. The system of coordinates is shown graphically in figure 7.4. We will derive here the form for \hat{l}_z , with a view

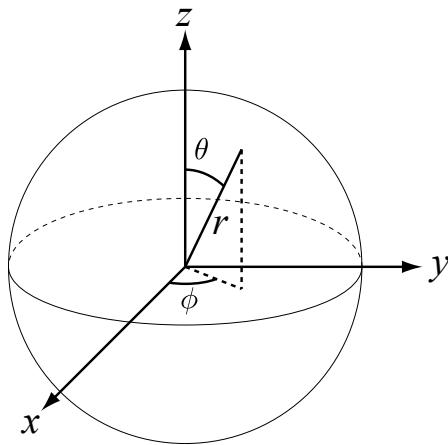


Figure 7.4: Sketch of angles chosen to define Spherical coordinates.

that the other operators can be derived in a similar way. Explicitly writing \hat{l}_z in terms of differential operators on x and y , we find

$$\hat{l}_z = -i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) . \quad (7.26)$$

Our next task is to find the partial derivatives in terms of r, θ, ϕ , for which we use the chain rule. The results (you should go and verify this) are

$$\frac{\partial}{\partial x} = \sin \theta \cos \phi \frac{\partial}{\partial r} + \frac{\cos \theta \cos \phi}{r} \frac{\partial}{\partial \theta} - \frac{\sin \phi}{r \sin \theta} \frac{\partial}{\partial \phi} \quad (7.27)$$

$$\frac{\partial}{\partial y} = \sin \theta \sin \phi \frac{\partial}{\partial r} + \frac{\cos \theta \sin \phi}{r} \frac{\partial}{\partial \theta} + \frac{\cos^2 \phi}{r \sin \theta \cos \phi} \frac{\partial}{\partial \phi} , \quad (7.28)$$

from which it takes simple trigonometric algebra to arrive at the final form for \hat{l}_z

$$\hat{l}_z = -i\hbar \frac{\partial}{\partial \phi} . \quad (7.29)$$

Eigenfunctions for this operator can be found from

$$-i\hbar \frac{\partial \psi_m}{\partial \phi} = \hbar m \psi_m \quad (7.30)$$

giving

$$\psi_m(\phi) = A e^{im\phi} \quad (7.31)$$

There is a critical requirement in quantum mechanics that the wave function must only have a single value at any point in space. This restricts the permissible values of m to integers, since on rotating the co-ordinate system by an angle of 2π it returns to its original position. Note how the symmetry here imposes a boundary condition which leads to quantization - we merely said that the system must be rotationally symmetric and we obtained quantized values!

The other orbital angular momentum operators can also be expanded in terms of spherical co-ordinates. We will look at the solutions later in the lecture. However for now let us step back and have a look at other properties of these operators, namely their commutation relations. Because we know already the commutation relations between x, \hat{p}_x etc., we don't need the co-ordinate forms of the operators \hat{l}_x, \hat{l}_y etc. to evaluate their commutation relations. One example is for \hat{l}_x and \hat{l}_y , where we find

$$[\hat{l}_x, \hat{l}_y] = [y\hat{p}_z - z\hat{p}_y, z\hat{p}_x - x\hat{p}_z] = ([y\hat{p}_z, z\hat{p}_x] - [y\hat{p}_z, x\hat{p}_z] - [z\hat{p}_y, z\hat{p}_x] + [z\hat{p}_y, x\hat{p}_z]) \quad (7.32)$$

In order to evaluate this, we note that the only operators acting on the same co-ordinate in this expression are for z . Evaluating these commutators using the result $[z, \hat{p}_z] = -i\hbar$ that we found the other day, we find that

$$[\hat{l}_x, \hat{l}_y] = -i\hbar(y\hat{p}_x - x\hat{p}_y) = i\hbar\hat{l}_z \quad . \quad (7.33)$$

The other relations are

$$[\hat{l}_x, \hat{l}_z] = -i\hbar\hat{l}_y \quad (7.34)$$

$$[\hat{l}_y, \hat{l}_z] = i\hbar\hat{l}_x \quad (7.35)$$

It is worth noting at this point that the commutation relations here imply that

$$\begin{aligned} [\hat{l}^2, \hat{l}_z] &= [\hat{l}_x^2 + \hat{l}_y^2 + \hat{l}_z^2, \hat{l}_z] \\ &= [\hat{l}_x^2, \hat{l}_z] + [\hat{l}_y^2, \hat{l}_z] \\ &= \hat{l}_x[\hat{l}_x, \hat{l}_z] + [\hat{l}_x, \hat{l}_z]\hat{l}_x + \hat{l}_y[\hat{l}_y, \hat{l}_z] + [\hat{l}_y, \hat{l}_z]\hat{l}_y \\ &= -i\hbar(\hat{l}_x\hat{l}_y + \hat{l}_y\hat{l}_x) + i\hbar(\hat{l}_x\hat{l}_y + \hat{l}_y\hat{l}_x) \\ &= 0 \end{aligned} \quad (7.36)$$

This result tells us that we can have a complete set of states which are simultaneously eigenstates of \hat{l}^2 and \hat{l}_z .

We can form other operators by linear combinations of $\hat{l}_x, \hat{l}_y, \hat{l}_z$, and it turns out that these can be very useful. In particular \hat{l}_+ and \hat{l}_- , defined as

$$\hat{l}_+ = \hat{l}_x + i\hat{l}_y \quad (7.37)$$

$$\hat{l}_- = \hat{l}_x - i\hat{l}_y \quad (7.38)$$

can be seen to play the role of raising and lowering operators in an analogous fashion to \hat{a}^\dagger and \hat{a} in the case of the harmonic oscillator. We can see this by applying \hat{l}_+ to the eigenvalue

equation for \hat{l}_z and making use of the commutation relation $[\hat{l}_+, \hat{l}_z] = -\hbar\hat{l}_+$ which follows from equations 7.34 and 7.35

$$\hat{l}_+ \hat{l}_z \psi_m = \hat{l}_+ m \hbar \psi_m \quad (7.39)$$

$$(\hat{l}_z - \hbar) \hat{l}_+ \psi_m = m \hbar \hat{l}_+ \psi_m \quad (7.40)$$

On defining $\psi_k = \hat{l}_+ \psi_m$, this gives us a new eigenvalue equation

$$\hat{l}_z \psi_k = (m + 1) \hbar \psi_k \quad (7.41)$$

which motivates us to write $k = m + 1$. We can see then that this operator can lead us to a set of angular momentum states with a ladder of angular momentum projections along z . In a similar manner, $(\hat{l}_- \psi_m) = \psi_k$ produces an eigenstate with $k = m - 1$.

We are now ready to consider the \hat{l}^2 operator again. As usual with quantum mechanics, we are interested in finding the eigenvectors and eigenvalues of the operator. We will try and write it in terms of the operators we have just been considering, \hat{l}_+ , \hat{l}_- and \hat{l}_z . We first note that

$$\hat{l}_+ \hat{l}_- = \hat{l}_x^2 + \hat{l}_y^2 + i(\hat{l}_y \hat{l}_x - \hat{l}_x \hat{l}_y) = \hat{l}_x^2 + \hat{l}_y^2 + \hbar \hat{l}_z \quad (7.42)$$

$$\hat{l}_- \hat{l}_+ = \hat{l}_x^2 + \hat{l}_y^2 - i(\hat{l}_y \hat{l}_x - \hat{l}_x \hat{l}_y) = \hat{l}_x^2 + \hat{l}_y^2 - \hbar \hat{l}_z \quad (7.43)$$

and then use this result to find the eigenvalues of

$$(\hat{l}_x^2 + \hat{l}_y^2 + \hat{l}_z^2) \psi_l = \omega_l \psi_l \quad (7.44)$$

On inserting 7.42, we find

$$(\hat{l}_+ \hat{l}_- - \hbar \hat{l}_z + \hat{l}_z^2) \psi_l = \omega_l \psi_l \quad (7.45)$$

which indicates that

$$\hat{l}_+ \hat{l}_- \psi_l = (\omega_l + \hbar \hat{l}_z - \hat{l}_z^2) \psi_l \quad (7.46)$$

If we choose ψ_l to be simultaneously an eigenstate of \hat{l}_z with eigenvalue $m\hbar$, we find

$$\hat{l}_+ \hat{l}_- \psi_{l,m} = (\omega_l - \hbar^2 m(m-1)) \psi_{l,m} \quad (7.47)$$

A similar analysis leads to

$$\hat{l}_- \hat{l}_+ \psi_{l,k} = (\omega_l - \hbar^2 k(k+1)) \psi_{l,k} \quad (7.48)$$

Let us suppose that there is a maximum allowable quantum number k_{\max} . Then the left hand side of this equation tells us that the right hand side must be zero, in which case for non-zero $\psi_{l,k_{\max}}$

$$\omega_l = \hbar^2 k_{\max}(k_{\max} + 1). \quad (7.49)$$

Similarly, let us suppose there is also a minimum eigenvalue for \hat{l}_z , which we will call m_{\min} . Using equation 7.47, we find

$$\omega_l = \hbar^2 m_{\min}(m_{\min} - 1) \quad (7.50)$$

Since ω_l does not depend on m , we can equate these two equations

$$m_{\min}(m_{\min} - 1) = k_{\max}(k_{\max} + 1) \quad (7.51)$$

which is satisfied for $m_{\min} = -k_{\max}$. We thus choose to call $k_{\max} = l$, the quantum number for the total angular momentum operator squared \hat{l}^2 . The eigenvalue equation for this operator then looks like

$$\hat{l}^2 \psi_{l,m} = \hbar^2 l(l+1) \psi_{l,m} \quad (7.52)$$

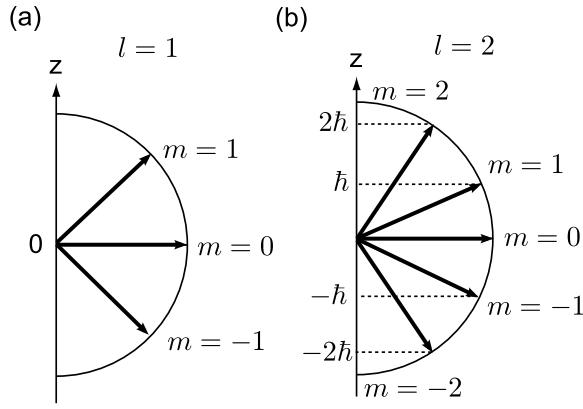


Figure 7.5: For every value of total angular momentum eigenvalue l there are $2l + 1$ different values of m_l ranging from $-l$ to l . Shown are vectors which have $|\hat{l}^2| = \hbar\sqrt{l(l+1)}$ and different values of m_l . Since m_l is the eigenvalue of the \hat{l}_z operator, these quantized values correspond to a discrete set of projections of the angular momentum onto the z axis.

for all eigenstates of the \hat{l}_z operator. The eigenvalues of \hat{l}_z must lie between l and $-l$, and must take integer values. This implies that there are $2l + 1$ eigenstates of \hat{l}_z for every value of l . Figure 7.5 tries to illustrate this principle graphically.

So now we know the eigenvalues we should get, let us apply ourselves to getting hold of the eigenstates. Since the \hat{l}_z operator acts purely on ϕ , we can write the solution as a product of two functions

$$\psi_{l,m} = H_{l,m}(\theta)\Phi_m(\phi) \quad (7.53)$$

where as we saw earlier

$$\Phi_m(\phi) = e^{im\phi} \quad (7.54)$$

For the maximal value of l , we know that

$$\hat{l}_+ H_{l,l}(\theta) e^{il\phi} = 0 \quad (7.55)$$

Using the chain rule (you will do this in the problem sets), we can derive expressions for \hat{l}_x and \hat{l}_y , and thus find

$$\hat{l}_+ = \hbar e^{i\phi} \left(\frac{\partial}{\partial \theta} + i \cot(\theta) \frac{\partial}{\partial \phi} \right) \quad (7.56)$$

$$\hat{l}_- = \hbar e^{-i\phi} \left(-\frac{\partial}{\partial \theta} + i \cot(\theta) \frac{\partial}{\partial \phi} \right) . \quad (7.57)$$

As a result, equation 7.55 gives a differential equation in θ :

$$\frac{\partial H_{l,l}}{\partial \theta} = l \cot(\theta) H_{l,l} . \quad (7.58)$$

We can see that this is of the form

$$\frac{dH_{l,l}}{H_{l,l}} = l \frac{du}{u} \quad (7.59)$$

where $u = \sin \theta$. Therefore, the solution is

$$H_{l,l} = K \sin^l \theta \quad (7.60)$$

where K is an integration constant. This can be found by normalization, where this is performed over all angles. The normalization thus gives

$$1 = \int_0^\pi \int_0^{2\pi} K^2 (\sin \theta)^{2l} \sin \theta d\theta d\phi = 4\pi K^2 \int_0^\pi (\sin \theta)^{2l+1} d\theta \quad (7.61)$$

This is not the easiest integral, so I will just present the result, which is

$$1 = 4\pi K^2 \frac{(l!)^2 2^{2l}}{(2l+1)!} \quad (7.62)$$

and thus

$$K = \frac{1}{\sqrt{4\pi}} \frac{\sqrt{(2l+1)!}}{l! 2^l} \quad (7.63)$$

and we therefore find

$$H_{l,l}(\theta) = \frac{1}{\sqrt{4\pi}} \frac{\sqrt{(2l+1)!}}{l! 2^l} \sin^l \theta \quad (7.64)$$

and, from equation 7.55

$$\psi_{l,l}(\theta, \phi) = \frac{1}{\sqrt{4\pi}} \frac{\sqrt{(2l+1)!}}{l! 2^l} \sin^l \theta e^{il\phi} . \quad (7.65)$$

The wavefunctions for other values of m can now be generated by repetitive application of \hat{l}_- . The set of polynomial solutions which is produced by this method gives the generalized Legendre polynomials which are commonly written as $Y_{l,m}(\theta, \phi)$. The first few of these polynomials are given below.

$$l = 0, \quad Y_{0,0} = \frac{1}{\sqrt{4\pi}} \quad (7.66)$$

$$l = 1, m = 1, \quad Y_{1,1} = \sqrt{\frac{3}{8\pi}} \sin \theta e^{i\phi} \quad (7.67)$$

$$l = 1, m = 0, \quad Y_{1,0} = \sqrt{\frac{3}{4\pi}} \cos \theta \quad (7.68)$$

$$l = 1, m = -1, \quad Y_{1,-1} = \sqrt{\frac{3}{8\pi}} \sin \theta e^{-i\phi} \quad (7.69)$$

$$l = 2, m = 2, \quad Y_{2,2} = \frac{1}{4} \sqrt{\frac{15}{8\pi}} \sin^2 \theta e^{2i\phi} . \quad (7.70)$$

Some plots of $|Y_{l,m}(\theta, \phi)|^2$ are plotted as a function of θ and ϕ in figure 7.6. You should remember if you are looking to count nodes of the $Y_{l,m}$ that these are given by zero crossings of the wavefunction itself. In the angular case, this is more confusing than for the square well we looked at earlier. The way to do this is to look at the function, and ask how many times it flips in sign as you vary θ between 0 and π and then as you vary ϕ between 0 and 2π . Bearing this in mind, we note that these are the lowest eigenstates of the energy operator for angular momentum and there is the same node counting behavior that we observed in previous problems.

Since the \hat{l}^2 operator describes the angular behavior in the Hamiltonian of the Hydrogen atom, it makes sense that the $Y_{l,m}(\theta, \phi)$ describe the angular parts of the wavefunction for this problem. This is the motivation behind our separation of variables using the function $Y(\theta, \phi)$, though now we see that we get eigenfunctions which are also labeled by the eigenvalues of total and z angular momentum l and m .

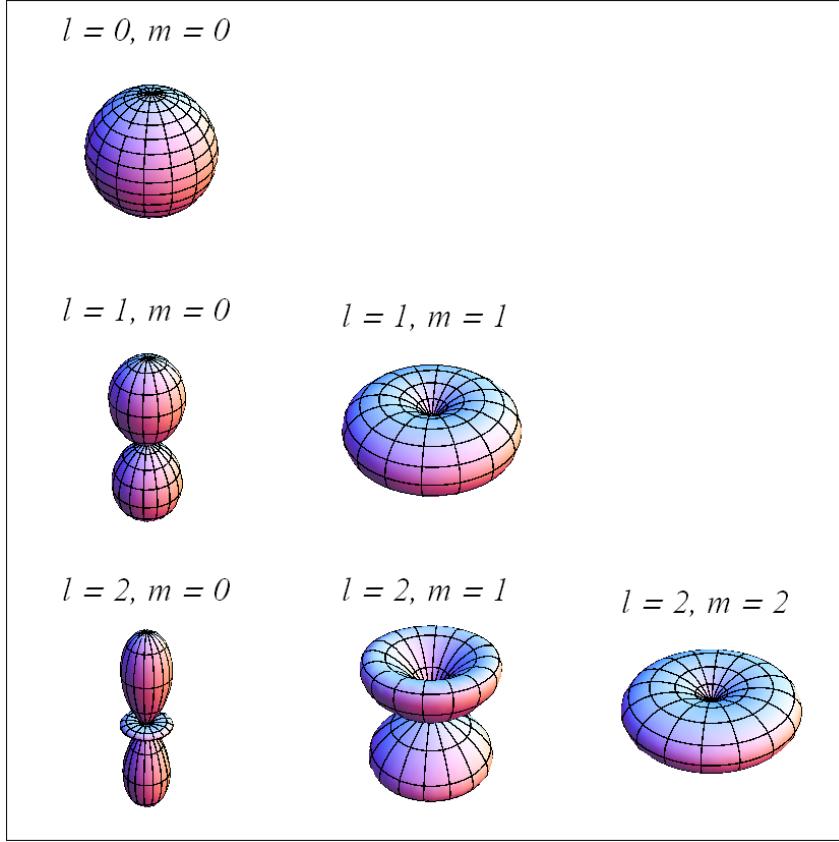


Figure 7.6: The first few spherical Harmonics $|Y_{l,m}|^2$ given as a distance from the center as a function of θ, ϕ . The polar axis corresponds to $\phi = 0$.

7.2.3 The radial solutions

In the previous section we found that the eigenvalues for the angular momentum satisfy the equation

$$\hat{L}^2 Y_{l,m}(\theta) \Phi_m(\phi) = \hbar^2 l(l+1) Y_{l,m}(\theta) \Phi_m(\phi) \quad (7.71)$$

With this in mind, we now re-visit equation 7.21 and look for solutions to the radial part. The radial component to the wavefunction will satisfy

$$\frac{-\hbar^2}{2mR} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} R \right) + \frac{\hbar^2 l(l+1)}{2m} = \left(\frac{e^2 r}{4\pi\epsilon_0} + E r^2 \right) \quad (7.72)$$

Multiplying everything by $-2mR/(\hbar^2 r^2)$ gives us a simpler form

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} R \right) - \frac{l(l+1)}{r^2} R = \left(-\frac{2me^2}{4\pi\epsilon_0\hbar^2 r} - \frac{2mE}{\hbar^2} \right) R \quad (7.73)$$

and on expanding the differential term we get

$$\frac{\partial^2 R}{\partial r^2} + \frac{2}{r} \frac{\partial R}{\partial r} - \frac{l(l+1)}{r^2} R = \left(-\frac{2me^2}{4\pi\epsilon_0\hbar^2 r} - \frac{2mE}{\hbar^2} \right) R \quad (7.74)$$

This looks complicated enough that we'll need a power series approach to solve it. The method we will use is due to Frobenius. It involves inserting a trial solution which is of the form of a power series

$$R(r) = \sum_j a_j r^j . \quad (7.75)$$

If this solution is to be valid for all r , then we require that the coefficients of r^j from one side of the equation equal the coefficients of r^j from the other side. This allows us to obtain a set of relations between the a_j . Typically this gives a power series solution to the equation we are trying to solve. It is always worth checking that this power series does not increase infinitely, and in some cases we can recognize the obtained power series as a combination of known functions, such as sin and cos.

Before moving to a series solution, it is a good idea to simplify the form of the solution first. For this reason, let us first look at equation 7.74 in the limit where $r \rightarrow \infty$. For this condition, equation 7.74 becomes

$$\frac{\partial^2 R}{\partial r^2} = -\frac{2mE}{\hbar^2} R \quad (7.76)$$

Remember that we are looking for solutions where the electron is bound, ie. where the negative contribution from the potential energy is larger in magnitude than the kinetic energy. The total energy will therefore be negative. Thus we see that in the large r limit the radial solutions should look like

$$R \propto e^{-\kappa r} , \quad (7.77)$$

where $\kappa = \sqrt{2m|E|}/\hbar$. This implies that we might be sensible to look for solutions for all r of the form

$$R = S(r)e^{-\kappa r} \quad (7.78)$$

where $S(r)$ is a solution to

$$\frac{\partial^2 S}{\partial r^2} + \left(\frac{2}{r} - 2\kappa \right) \frac{\partial S}{\partial r} + \left[\frac{-2\kappa}{r} - \frac{l(l+1)}{r^2} + \left(\frac{2me^2}{4\pi\epsilon_0\hbar^2 r} \right) \right] S = 0 \quad (7.79)$$

Let us also simplify the equation by reducing the number of constants. The collection of constants on the right hand side is related to the Bohr radius $1/a_0 = me^2/(4\pi\epsilon_0\hbar^2)$ (equation 7.13), and we can also simplify the equation by making the substitution $\rho = 2\kappa r$. Since this means that $d\rho = 2\kappa dr$, we find a somewhat simplified form

$$\frac{\partial^2 S}{\partial \rho^2} + \left(\frac{2}{\rho} - 1 \right) \frac{\partial S}{\partial \rho} + \left[\left(\frac{1}{\kappa a_0} - 1 \right) \frac{1}{\rho} - \frac{l(l+1)}{\rho^2} \right] S = 0 . \quad (7.80)$$

We now insert a general polynomial solution

$$S(\rho) = \sum_{i=0}^{\infty} c_i \rho^{i+j} \quad (7.81)$$

where the c_i and j are still to be determined. We note that for this form

$$\begin{aligned} \frac{\partial S}{\partial \rho} &= \sum_{i=0}^{\infty} c_i (i+j) \rho^{i+j-1} \\ \frac{\partial^2 S}{\partial \rho^2} &= \sum_{i=0}^{\infty} c_i (i+j)(i+j-1) \rho^{i+j-2} \end{aligned} \quad (7.82)$$

so equation 7.80 is now written as

$$\begin{aligned} 0 &= \sum_{i=0}^{\infty} c_i (i+j)(i+j-1) \rho^{i+j-2} + \sum_{i=0}^{\infty} 2c_i (i+j) \rho^{i+j-2} - \sum_{i=0}^{\infty} c_i (i+j) \rho^{i+j-1} \\ &+ \left(\frac{1}{\kappa a_0} - 1 \right) \sum_{i=0}^{\infty} c_i \rho^{i+j-1} - l(l+1) \sum_{i=0}^{\infty} c_i \rho^{i+j-2} \end{aligned} \quad (7.83)$$

Since the lowest value of i is zero, the lowest power of ρ which is found in this expression is ρ^{j-2} , for which we find

$$\begin{aligned} 0 &= c_0 j(j-1) + c_0 2j - c_0 l(l+1) \\ 0 &= j(j+1) - l(l+1) . \end{aligned} \quad (7.84)$$

There are only two possible solutions to this equation. Either $j = l$, or $j = -l - 1$. If we look at our expansion for $S(\rho)$, we quickly realize that the latter leads to a solution which is infinite at $\rho = 0$. This is clearly not a satisfactory solution to the Schrödinger equation, so we end up choosing $j = l$.

Looking at higher orders of ρ , and again equating coefficients (now of ρ^{i+j-2}) we find

$$c_i(i+l)(i+l-1) + 2c_i(i+l) - c_{i-1}(i-1+l) + \left(\frac{1}{\kappa a_0} - 1 \right) c_{i-1} - l(l+1)c_i = 0 \quad (7.85)$$

which gives a recursion relation

$$c_i = \frac{i+l-\frac{1}{\kappa a_0}}{(i+l)(i+l+1)-l(l+1)} c_{i-1} . \quad (7.86)$$

There are two possible solutions which can be derived from this equation. Either the solution looks like a power series which contains an infinite number of terms, or it terminates at some point. We must remember that the wavefunction must be finite, so if the infinite series does not converge then it will not produce a valid solution. It turns out that this is the case for the recursion relation above. To produce a finite solution for R from this series we must make sure that it terminates, that is, for some value of i the numerator goes to zero. This can only happen in the expression above if

$$i+l-\frac{1}{\kappa a_0} \quad (7.87)$$

vanishes for some i . Since i and l are integers, this implies that we should identify $1/(\kappa a_0)$ with an integer, which we will call n . This condition places a restriction on the allowed energy levels of the hydrogen atom, since

$$\left(\frac{1}{\kappa a_0} \right)^2 = \frac{1}{a_0^2} \left(\frac{\hbar^2}{2m(-E_n)} \right) = n^2 \quad (7.88)$$

and thus

$$E_n = -\frac{1}{n^2} \left(\frac{me^4}{32\pi^2\epsilon_0^2\hbar^2} \right) = -\frac{1}{n^2} \frac{mc^2\alpha^2}{2} \quad (7.89)$$

where on the right hand side I have introduced the fine structure constant α .

We can see from the above that the radial wave functions for the hydrogen atom depend on n , which we call the *principal quantum number* of the energy level, and also on l , which is the quantum number for the *orbital angular momentum* of the electron. For each different value of n, l , we will obtain a different polynomial $S(\rho)$ (and thus $R(r)$), because the coefficients are determined by equation 7.86. For each different value of n , the exponential factor $e^{-\kappa r}$ will differ due to the n -dependence of κ .

Let us now go back to find the radial wave functions. The co-efficients of the polynomials in the Frobenius expansion are given by

$$c_i = \prod_{q=1}^i \frac{q+l-n}{(q+l)(q+l+1)-l(l+1)} c_0 . \quad (7.90)$$

It is possible to write this in a more compact form. Let us expand the denominator, and then simplify it

$$(q+l)(q+l+1) - l(l+1) = q(q+l+1) + ql + l(l+1) - l(l+1) = q(q+2l+1) \quad (7.91)$$

which gives us

$$c_i = \prod_{q=1}^i \frac{q+l-n}{q(q+2l+1)} c_0 = (-1)^i \frac{(n-l-i)!}{i!(i+2l+1)!} c_0. \quad (7.92)$$

This line holds for $i < n-l$. Remember that all other $c_i = 0$ for $i \geq n-l$. The solutions we need to solve equation 7.80 are therefore given by

$$S(\rho) = \rho^l c_0 \left[\sum_{i=0}^{n-l-1} (-1)^i \frac{(n-l-i)!}{i!(i+2l+1)!} \rho^i \right]. \quad (7.93)$$

Polynomials of the form which is given within the square brackets are well known in mathematics, they are the generalized Laguerre polynomials. It is rare to see them written in the form above, but be aware that there are two standard definitions that get used in different texts. The one I favor is

$$L_n^\alpha(x) = \frac{x^{-\alpha} e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^{n+\alpha}) \quad (7.94)$$

and the other is

$$L_k^\beta(x) = \frac{d^\beta}{dx^\beta} \left(e^x \frac{d^k}{dx^k} (e^{-x} x^k) \right). \quad (7.95)$$

The two definitions require different combinations of n, α and k, β to produce the same result. Totally stupid I know, but that is life. Mathematica uses the first definition, which is what I shall use here. Haken and Wolf use the second definition. The crazy thing about this is that in no place I looked do people declare that there are two usages of the same symbol L . This is nuts!

Using the first definition,

$$S(\rho) = c_0 \rho^l L_{n-l-1}^{2l+1}(\rho) \quad (7.96)$$

where c_0 can be found by normalization. Transforming back to $R(r)$, we find that

$$R(r) = S(\rho) e^{-\rho/2} = S(2\kappa r) e^{-\kappa r}, \quad (7.97)$$

which along with $\kappa = 1/(na_0)$ gives

$$R_{n,l}(r) = S_{n,l}(2r/(na_0)) e^{-r/(na_0)} \quad (7.98)$$

$$= C r^l L_{n-l-1}^{2l+1}(2r/(na_0)) e^{-r/(na_0)} \quad (7.99)$$

where $C \neq c_0$ can be determined by normalization. It is worth noting that this is not the wave function - the wave function includes also the angular parts, which we derived above.

The first few radial wave functions are

$$R_{1,0}(r) = \frac{2}{a_0^{3/2}} e^{-r/a_0} \quad (7.100)$$

$$R_{2,0}(r) = \frac{1}{(2a_0)^{3/2}} e^{-r/(2a_0)} \left(2 - \frac{r}{a_0} \right) \quad (7.101)$$

$$R_{2,1}(r) = \frac{1}{\sqrt{3}(2a_0)^{3/2}} \left(\frac{2r}{2a_0} \right) e^{-r/(2a_0)} \quad (7.102)$$

where the normalization has been performed such that $\int_0^\infty r^2|R(r)|^2dr = 1$. Note that we should ensure that the normalization for the angular and radial parts are consistent; in the end, we care that the total wave function is normalized. This requires

$$\int_0^\infty r^2|R_{n,l}(r)|^2dr \int_0^\pi \int_0^{2\pi} |Y_{l,m}(\theta, \phi)|^2 \sin(\theta)d\theta d\phi = 1 \quad (7.103)$$

Since my normalization of the angular part of the wave functions was such that

$$\int_0^\pi \int_0^{2\pi} |Y_{l,m}(\theta, \phi)|^2 \sin(\theta)d\theta d\phi = 1 \quad (7.104)$$

my radial normalization is

$$\int_0^\infty r^2|R_{n,l}(r)|^2dr = 1 \quad . \quad (7.105)$$

Note that the probability to find the particle at radial distance between r and $r + dr$ is given by

$$r^2|R_{n,l}(r)|^2dr \int_0^\pi \int_0^{2\pi} |Y_{l,m}(\theta, \phi)|^2 \sin(\theta)d\theta d\phi = r^2|R_{n,l}(r)|^2dr \quad . \quad (7.106)$$

This is a very important point which is often missed - please don't fall into this trap.

The radial wave functions are plotted in figure 7.7 and $r^2R(r)$ in figure 7.8

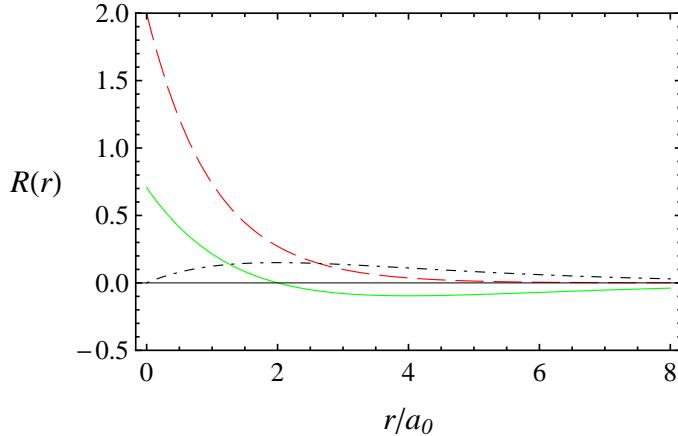


Figure 7.7: The lowest energy radial components of the wave function $R_{n,l}$ for the Hydrogen atom. $n = 1, l = 0$: red, dashed. $n = 2, l = 0$: green, solid. $n = 2, l = 1$, black, dot-dashed. Note that the number of nodes is $n - l - 1$. For convenience, a_0 has been set to 1 (otherwise the numbers are huge).

Number of nodes, radial wave functions

At this point, I want to add a note about the number of times the radial wave function $R_{n,l}$ will cross zero $n - l - 1$ times. The reason is simple. As for the square well, the ground state does not cross zero, but the n 'th excited state crosses $n - 1$ times. However we should also remember that the *angular* part of the wave function can have nodes; in fact for angular momentum quantum number l it will have l nodes. So in order that the *total* wave function has $n - 1$ nodes, the *radial* wave function will have $n - l - 1$ nodes.

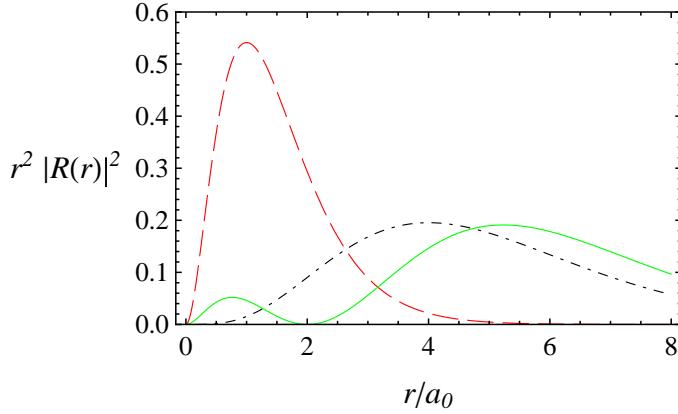


Figure 7.8: Probability $r^2|R_{n,l}(r)|^2$ of finding the atom between r and $r + dr$. $n = 1, l = 0$: red, dashed. $n = 2, l = 0$: green, solid. $n = 2, l = 1$, black, dot-dashed.

7.2.4 Degeneracy of the Hydrogen wave functions

The eigenvalues for energy for the Hydrogen atom are dependent on the single quantum number n . However for each of these values, we can have different combinations of l and m , corresponding to different angular wave functions. How many eigenstates exist for each n ?

We already noted that for every value of l , there are $2l + 1$ values of m , since m takes every integer value between l and -1 (including zero). The remaining question is how many values of l we are allowed for each value of n . To count these, let us return to equation 7.87. We note that the value of i for which the series will cut off is

$$i_{\max} = n - l . \quad (7.107)$$

The series starts at c_0 , thus from equation 7.86 we can see that the minimum value of i is 1 in this recursion relation. Therefore

$$n - l \geq 1 \quad (7.108)$$

and we can never have l as large as n . Therefore for the level with principal quantum number n , we have degeneracy

$$\sum_{l=0}^{n-1} (2l + 1) = n^2 \quad (7.109)$$

7.3 First approximations

For all the problems we have encountered until now, we have written out the full Hamiltonian and tried to solve it exactly in order to find the eigenfunctions and their associated eigenvalues. This is possible in these cases because we considered simple models, but in general physics is far too complex to solve exactly. Instead we have to resort to approximate methods. The most common of these is called *perturbation theory*, though other methods include *variational* approaches. In order to keep things relatively simple, here we will describe perturbation theory, which will allow us to examine a variety of effects which are exhibited in atoms due to the presence of external fields (electric and magnetic), or more complex interactions between particles.

We will consider a situation in which we have a dominant Hamiltonian \hat{H}_0 for which we can solve the TISE to find a set of eigenstates ψ_n and energy eigenvalues $E_n^{(0)}$, and a smaller contribution which adds a term $\delta\hat{H}$ which makes solving the TISE difficult. The total Hamiltonian is therefore given by

$$\hat{H} = \hat{H}_0 + \delta\hat{H} . \quad (7.110)$$

If the contribution due to $\delta\hat{H}$ is small compared to that due to \hat{H}_0 , then we would expect that:

1. The energy eigenvalues E_n do not change from $E_n^{(0)}$ by much.
2. The modified energy eigenfunctions $\psi_n^{(1)}$ are close to the original ψ_n .

If this is true (it is best to check this after applying the method), then we can write the new energy eigenvalues as

$$E_n = E_n^{(0)} + \delta E_n . \quad (7.111)$$

The new energy eigenfunctions can also be written using the old energy eigenfunctions. As we know, since the eigenfunctions of the original Hamiltonian ψ_k form a complete set, it is always possible to write

$$\psi_n^{(1)} = \sum_k a_k \psi_k . \quad (7.112)$$

for any wavefunction. We will make the approximation that this is dominated by the term with $k = n$, therefore we write

$$\psi_n^{(1)} = \psi_n + \delta\psi_n \quad (7.113)$$

where we note that since $\delta\psi_n$ is made up of all eigenfunctions other than ψ_n it is orthogonal to ψ_n (the normalization of this modified eigenfunction is not correct, but this can typically be dealt with after performing perturbation theory). Stated mathematically

$$\delta\psi_n = \sum_{k \neq n} c_k \psi_k \quad (7.114)$$

and thus

$$\int \psi_n^* \delta\psi_n dx = \sum_{k \neq n} c_k \int \psi_n^* \psi_k dx = 0 . \quad (7.115)$$

Our task is to find the values of the two adjustments δE_n and $\delta\psi_n$. We start by writing down the TISE for this problem.

$$(\hat{H}_0 + \delta\hat{H})(\psi_n + \delta\psi_n) = (E_n^{(0)} + \delta E_n)(\psi_n + \delta\psi_n) . \quad (7.116)$$

Now let us group terms which are on the same size scale. Remember that \hat{H}_0 , $E_n^{(0)}$ and ψ_n are considered as large compared to their adjustments $\delta\hat{H}$, δE_n and $\delta\psi_n$. Therefore keeping only terms which have these components we get

$$\hat{H}_0\psi_n = E_n^{(0)}\psi_n . \quad (7.117)$$

This is what we have solved in order to find the energy eigenstates of \hat{H}_0 , namely the unperturbed TISE. Now let us look at terms which are the next order of smallness. These will each involve one of the adjustment terms, but only one. We find

$$\delta\hat{H}\psi_n + \hat{H}_0\delta\psi_n = E_n^{(0)}\delta\psi_n + \delta E_n\psi_n . \quad (7.118)$$

Now we use our usual quantum mechanics trick, but in two different ways. To find the energy shift, we multiply through by ψ_n^* and integrate over x , making use of the orthogonality of the eigenfunctions. We find

$$\begin{aligned} \int \psi_n^* \delta\hat{H}\psi_n dx &= \delta E_n \int \psi_n^* \psi_n dx \\ &= \delta E_n \end{aligned} \quad (7.119)$$

and thus the energy shift of the n th level is

$$\delta E_n = \int \psi_n^* \delta\hat{H}\psi_n dx \quad (7.120)$$

To find the shift in the wavefunction, we need to find the set of coefficients c_k . To obtain these, we multiply through by ψ_m^* , where $m \neq n$, and integrate. This gives

$$\begin{aligned} \int \psi_m^* \delta\hat{H}\psi_n dx + \int \psi_m^* \hat{H}_0 \sum_{k \neq n} c_k \psi_k dx &= E_n^{(0)} \int \psi_m^* \sum_{k \neq n} c_k \psi_k dx + 0 \\ \int \psi_m^* \delta\hat{H}\psi_n dx + \int \psi_m^* \sum_{k \neq n} E_k^{(0)} c_k \psi_k dx &= E_n^{(0)} c_m \\ \int \psi_m^* \delta\hat{H}\psi_n dx + E_m^{(0)} c_m &= E_n^{(0)} c_m \end{aligned} \quad (7.121)$$

which when re-arranged gives a formula for c_m

$$c_m = \frac{\int \psi_m^* \delta\hat{H}\psi_n dx}{E_n^{(0)} - E_m^{(0)}} . \quad (7.122)$$

The shift in the wavefunction is then given by

$$\delta\psi_n = \sum_{m \neq n} c_m \psi_m . \quad (7.123)$$

These are the results of *first order perturbation theory*. You must always be careful when using them to check that the results really are small adjustments, otherwise it is not valid to use these methods.

We must also be very careful in the presence of degenerate states of the original Hamiltonian \hat{H}_0 , since these could cause equation 7.122 to become infinite for some of the c_m . This would happen if

$$\int \psi_m^* \delta\hat{H}\psi_n dx \neq 0 \quad (7.124)$$

for two states with $E_m^{(0)} = E_n^{(0)}$. The method for getting around this problem is simple, but the details of why it works I don't want to go into here (you'll see them if you do more

advanced quantum mechanics next year). The first thing to note is that when we have degenerate states, the states which are degenerate are often distinguished by eigenvalues of an operator other than the energy (in the case of Hydrogen these are the l and m quantum numbers). We can choose any superposition of these degenerate states labeled by different values of l and m and we still get a valid energy eigenstate. This gives us a lot of freedom, and it is this freedom which allows us to choose sets of states for which

$$\int \psi_m^* \delta \hat{H} \psi_n dx = 0 \text{ when } E_n^{(0)} = E_m^{(0)} . \quad (7.125)$$

At this point we will leave this issue alone, and return to it when we come to a potential problem case.

7.4 The Normal Zeeman effect

In solving for the eigenfunctions of the Hydrogen atom, we chose that our angular momentum eigenstates should be defined by the eigenstates of \hat{l}^2 and \hat{l}_z . Here the choice to use \hat{l}_z was purely a matter of convenience, because we defined our coordinate system in such a way that \hat{l}_z only operates through the azimuthal coordinate ϕ . We would have found an identical set of eigenfunctions with respect to the physics of the problem if we had chosen \hat{l}_y or \hat{l}_x as operator.

Now we will change the physics by adding a magnetic field to the problem. We will orient the magnetic field along the z axis. A magnetic field can only interact with magnetic dipoles, so we have to search in our problem for a source of a magnetic dipole.

7.4.1 The dipole moment of the orbiting electron

The dipole moment of a classical current loop is given by $IA\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector normal to the plane of the loop. In order to derive a dipole moment of a single electron in an atom, we will imagine that it orbits so fast that we average over this motion, and see a mean current enclosing a certain area. For an electron in a classical orbit with angular momentum \mathbf{l} , the velocity v , mass m_e and orbit radius r are related to \mathbf{l} by

$$\mathbf{l} = m_e \omega r^2 \hat{\mathbf{n}} . \quad (7.126)$$

The effective current due to the electron orbit is the charge per second passing a particular point. Since the electron travels $2\pi r$ with each orbit, it thus passes a particular point on the orbit every $v/(2\pi r)$ seconds. The time averaged current is thus

$$I = -\frac{e\omega r}{2\pi r} \quad (7.127)$$

where e is the unit of charge. Since the area of the loop is $A = \pi r^2$, we have that the dipole moment of the electron is

$$IA\hat{\mathbf{n}} = -\frac{e\omega}{2\pi} \pi r^2 \hat{\mathbf{n}} = -\frac{e}{2m_e} \mathbf{l} \quad (7.128)$$

7.4.2 Interaction with the magnetic field

The energy of interaction between a dipole and a magnetic field is given by

$$U = -\boldsymbol{\mu} \cdot \mathbf{B} . \quad (7.129)$$

Let us consider that the magnetic field $\mathbf{B} = (0, 0, B)$, ie. its direction is along the z axis. Then the interaction between the classical dipole and the field looks like

$$U = \frac{e}{2m_e} \mathbf{l} \cdot \hat{\mathbf{z}} B = \frac{el_z}{2m_e} B \quad (7.130)$$

where l_z is the z component of the angular momentum. To move to the quantum mechanical picture, we insert the operator for \mathbf{l} or l_z , which are our operators $\hat{\mathbf{l}} = (\hat{l}_x, \hat{l}_y, \hat{l}_z)$ and \hat{l}_z . The magnetic moment operator written in terms of the quantum mechanical operator is therefore

$$\hat{\boldsymbol{\mu}} = \frac{e}{2m_e} \hat{\mathbf{l}} \quad (7.131)$$

where you should note that the hat on $\boldsymbol{\mu}$ indicates an operator rather than a unit vector. To find the expectation value of the energy for any of the eigenfunctions, we must evaluate

equation 7.120 using

$$\delta\hat{H} = \frac{e\hat{l}_z}{2m_e}B \quad (7.132)$$

which gives

$$\begin{aligned}\delta E &= \frac{eB}{2m_e} \int \psi_{n,l,m}^* \hat{l}_z \psi_{n,l,m} d\mathbf{r} \\ &= \frac{eB}{2m_e} \int_0^\infty r^2 |R_{n,l}(r)|^2 dr \int_0^\pi \int_{-\pi}^\pi Y_{l,m}^*(\theta, \phi) \hat{l}_z Y_{l,m}(\theta, \phi) \sin(\theta) d\theta d\phi \quad (7.133)\end{aligned}$$

which since the $Y_{l,m}(\theta, \phi)$ is an eigenstate of \hat{l}_z with eigenvalue $\hbar m$ gives

$$\delta E = \frac{eB}{2m_e} \hbar m = \frac{e\hbar}{2m_e} m B = \mu_B m B \quad (7.134)$$

where μ_B is called the Bohr magneton. The states with different values of m are therefore split up in energy evenly spaced by $\mu_B B$. The degeneracy of the Hydrogen atom is partially removed. For each n , there are now $n - 1$ values of l , but the states with a given l will now be split up in energy into $2l + 1$ separate levels.

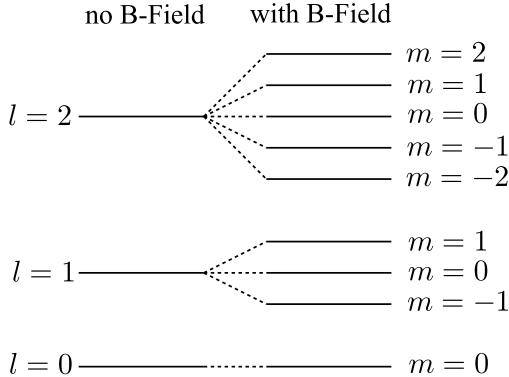


Figure 7.9: When a magnetic field is turned on, the degeneracy of the Hydrogen energy eigenstates is lifted and these split according to their quantum number m . Here the separation of energies at zero magnetic field is put in to illustrate the concept - our treatment of the hydrogen atom levels with the same l gives all these energy levels to be the same. In a few lectures time we will see an additional internal component to the hydrogen atom which actually splits these up.

7.4.3 Magnetic field aligned along x - the problem of degeneracy

Now let us illustrate the problem with degeneracy in quantum mechanics by considering a magnetic field aligned along the x axis. In this case, the interaction perturbation Hamiltonian would not be that written in equation 7.132, but would involve the \hat{l}_x operator

$$\delta\hat{H} = \frac{e\hat{l}_x}{2m_e}B \quad (7.135)$$

To evaluate this perturbation, it helps to write \hat{l}_x in terms of \hat{l}_+ and \hat{l}_- as

$$\hat{l}_x = \frac{1}{2}(\hat{l}_+ + \hat{l}_-) \quad (7.136)$$

We then see that the energy perturbation is

$$\delta E_x = \frac{eB}{4m_e} \int \psi_{n,l,m}^* (\hat{l}_+ + \hat{l}_-) \psi_{n,l,m} d\mathbf{r}$$

is zero, because the action of \hat{l}_+ and \hat{l}_- on $\psi_{n,l,m}$ is to alter the wavefunctions to $m \rightarrow m+1$ and $m \rightarrow m-1$ respectively. By orthogonality the integral will therefore equal zero. This seems completely bizarre - the hydrogen problem was originally completely spherically symmetric, so why should it matter in which direction I choose to put the magnetic field?

The problem lies not with the physics, but with our use of perturbation theory. If we looked for first order wavefunction shifts, we would find that for some of these the shift is infinite. So we need to have a new rule for perturbation theory in the presence of degeneracy. The rule is

- The set of energy eigenstates to use in degenerate perturbation theory should be those which are also eigenstates of operators which commute with both the original and perturbing Hamiltonian.

Can we make physical sense of this mathematical requirement? The answer is yes - or at least I feel happy that I can (whether you agree is another matter). The condition that the operator commutes with the Hamiltonian means that it shares eigenstates with the Hamiltonian operator, and thus that these states do not evolve in time (apart from the phase factor $e^{iEt/\hbar}$). These operators correspond to *constants of the motion*, or physical properties which are conserved. If you are looking for an operator to label your states, good choices are generally those quantities which are constant in classical motion. Once these are chosen, it is worth checking that their operators commute with both terms in the Hamiltonian.

In our example above, our eigenstates were defined in terms of the operators \hat{l}^2 and \hat{l}_z . For the case where the magnetic field pointed along z , these operators commute with the perturbation. However for the second case, we find that \hat{l}_z does not commute with the perturbation. In this case we should start by expressing the eigenstates of the original Hamiltonian in terms of eigenstates of \hat{l}^2 and \hat{l}_x , and then apply perturbation theory. Of course, in this case it is easier in this case to rotate the co-ordinate system (which has no effect on the unperturbed problem) such that the z axis lies along the magnetic field!

7.5 Optical transitions.

We now return to optical transitions, which we left a while ago now, while we were in the middle of the Bohr theory. The interactions between atoms and light is a complex field in its own right, much too complex to go into here. So we will not try to look at the full theory, but rather we shall try to see what type of transitions are allowed to happen, and which are not. To do this, we will delve briefly into the world of time-dependent Hamiltonians, a topic which we will come to again in the course, but also one which you will return to time and time again over the next few years.

A powerful tool in this area is the eigenfunction expansion. Remember that if we have a complete set of eigenfunctions for an operator, we can expand any wavefunction in terms of those operators. If our set was a set of energy eigenstates of a time *independent* Hamiltonian H_0 which are denoted by the subscript i , this means that I can write any wavefunction as

$$\psi(x) = \sum_i c_i \psi_i(x) \quad (7.137)$$

and all I need to do is find the complex coefficients c_i . Since these are energy eigenfunctions, I know that they evolve in time according to

$$\psi_i(x, t) = e^{-i\omega_i t} \psi_i(x, 0) \quad (7.138)$$

where $\omega_i \equiv E_i/\hbar$. Therefore at any instance in time, I can represent the wave function as

$$\psi(x, t) = \sum_i c_i e^{-i\omega_i t} \psi_i(x, 0) \quad (7.139)$$

where the c_i will be constants if the Hamiltonian is time-independent.

Now imagine that my total Hamiltonian consists of an additional piece $H_1(t)$. The time-dependent Schrödinger equation is then

$$[H_0 + H_1(t)] \psi(x, t) = i\hbar \frac{\partial}{\partial t} \psi(x, t) \quad (7.140)$$

Since the eigenfunctions of the time-independent part of the Hamiltonian $\psi_i(x, 0)$ are a complete set, I can still write the wavefunction at any instance in time in terms of these functions. What will change is the relative amplitudes c_i to be in each of the eigenstates. Thus the c_i now become time-dependent, and we write them as $c_i(t)$. Substituting equation 7.139 with time-dependent coefficients $c_i(t)$ into the Schrödinger equation gives

$$\begin{aligned} & \sum_i c_i(t) e^{-i\omega_i t} H_0 \psi_i(x, 0) + \sum_i c_i(t) e^{-i\omega_i t} (H_1(t) \psi_i(x, 0)) \\ &= \sum_i i\hbar \frac{\partial c_i(t)}{\partial t} e^{-i\omega_i t} \psi_i(x, 0) + \sum_i \hbar \omega_i c_i(t) e^{-i\omega_i t} \psi_i(x, 0) \end{aligned} \quad (7.141)$$

which using $H_0 \psi_i(x, 0) = E_i \psi_i(x, 0) = \hbar \omega_i \psi_i(x, 0)$ and canceling terms simplifies to

$$\sum_i c_i(t) e^{-i\omega_i t} H_1(t) \psi_i(x, 0) = \sum_i i\hbar \frac{\partial c_i(t)}{\partial t} e^{-i\omega_i t} \psi_i(x, 0) . \quad (7.142)$$

Now we again make use of one of the common tricks in quantum mechanical derivations. We know that eigenfunctions with different eigenvalues are orthogonal, so we multiply everything by $\psi_j^*(x, 0)$ and integrate over x

$$\begin{aligned} & \sum_i c_i(t) e^{-i\omega_i t} \int_{-\infty}^{\infty} \psi_j^*(x, 0) H_1(t) \psi_i(x, 0) dx \\ &= \sum_i i\hbar \frac{\partial c_i(t)}{\partial t} e^{-i\omega_i t} \int_{-\infty}^{\infty} \psi_j^*(x, 0) \psi_i(x, 0) dx \end{aligned} \quad (7.143)$$

which since the eigenfunctions are orthogonal for $i \neq j$ gives

$$\frac{\partial c_j(t)}{\partial t} e^{-i\omega_j t} = \frac{-i}{\hbar} \sum_i c_i(t) e^{-i\omega_i t} \int_{-\infty}^{\infty} \psi_j^*(x, 0) H_1(t) \psi_i(x, 0) dx \quad (7.144)$$

$$\frac{\partial c_j(t)}{\partial t} = \frac{-i}{\hbar} \sum_i c_i(t) e^{-i(\omega_i - \omega_j)t} \int_{-\infty}^{\infty} \psi_j^*(x, 0) H_1(t) \psi_i(x, 0) dx \quad (7.145)$$

Note that I have made no approximations whatsoever in deriving this form. This equation tells us that all (!!) we need to know to evaluate the time dependence under an arbitrary Hamiltonian is the set of what we call *matrix elements*

$$\int_{-\infty}^{\infty} \psi_j^*(x, 0) H_1(t) \psi_i(x, 0) dx . \quad (7.146)$$

They are called matrix elements because for each pair i, j we have a single number (at any one point in time), and this tells us all we need to know. Therefore at any instance in time we can write the evolution down as a matrix equation.

Of course, evaluating the expression above for every single time step is an impossible task in a system of any significant size. For our present purposes, we shan't do this, but instead we will note one simple thing. If we start in a state where $c_i(0) = 1$ and we want to have the system evolve into the state with amplitude $c_j(t) \neq 0$, we will need the matrix element given by equation 7.146 to be non-zero. If we find values which are zero, this transition is not allowed - it will not happen.

So how does this relate to optical transitions in the Hydrogen atom? Well it is important to consider the form of the Hamiltonian for the electromagnetic field. The full form is beyond the scope of the course, but we do know that the electric field propagates according to the form of a wave

$$\mathbf{E}(t) = \mathbf{E}(0) e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \quad (7.147)$$

this field will have an energy of interaction with the electron which is proportional to

$$U = \mathbf{E}(t) \cdot \mathbf{r} = \mathbf{E}(0) \cdot \mathbf{r} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \quad (7.148)$$

and therefore light polarized along the x direction will try to oscillate the electron along x , etc.. So long as the field is not too large, the electron will stay within the realms of the atom, which is small compared to typical wavelengths of light

$$\lambda \sim 600 \text{ nm}, r < 0.1 \text{ nm} \quad (7.149)$$

and therefore the $kr < 10^{-3}$ factor in the exponential can be neglected. The interesting matrix elements are then

$$\int \psi_{n,l,m}^*(r, \theta, \phi) (\mathbf{E}(0) \cdot \mathbf{r}) \psi_{n',l',m'}(r, \theta, \phi) r^2 dr \sin \theta d\theta d\phi \quad (7.150)$$

We must now relate $\mathbf{E}(0) \cdot \mathbf{r}$ to the coordinate system in which we derived the wavefunctions. Since the direction of $\mathbf{E}(0)$ is determined by the polarization of the light, we will consider the different polarization components of the light separately.

7.5.1 Linear polarization aligned with an external magnetic field

If the light is linearly polarized in the same direction as a magnetic field which points along z , then

$$\mathbf{E}(0) \cdot \mathbf{r} = E_0 z = E_0 r \cos \theta \quad (7.151)$$

where E_0 is the electric field amplitude. With this in mind, we see that if the integral

$$\int \psi_{n,l,m}^*(r, \theta, \phi) \psi_{n',l',m'}(r, \theta, \phi) r^3 dr \cos \theta \sin \theta d\theta d\phi \quad (7.152)$$

is non-zero then we can drive a transition. For what pairs of n, n' , l, l' and m, m' is this possible? Let us consider the integrals over r , θ and ϕ separately. We will deal with these in the reverse order, since the selection rule for r offers less intuition than those for θ and ϕ .

Selection rule on m

Remember that the wave functions $\psi_{n,l,m}(r, \theta, \phi)$ and $\psi_{n',l',m'}(r, \theta, \phi)$ are both of the form

$$\psi_{n,l,m}(r, \theta, \phi) = R(r)Y_{l,m}(\theta, \phi) \quad (7.153)$$

where

$$Y_{l,m} = N e^{im\phi} H_{l,m}(\theta) \quad (7.154)$$

with the respective values n, l, m or n', l', m' . They are dependent on ϕ only through the \hat{l}_z component of the wave function which is of the form $e^{im\phi}$ and $e^{im'\phi}$ respectively. Therefore we have that

$$\psi_{n,l,m}^*(r, \theta, \phi) \psi_{n',l',m'}(r, \theta, \phi) \propto e^{i(m'-m)\phi} . \quad (7.155)$$

We will integrate this function over ϕ , from $0 \rightarrow 2\pi$. Since both m' and m are integers, the integral will only be non-zero for $m' - m = 0$. This is our first *selection rule*. It tells us that we can only have transitions for which $m' = m$ if we use light for which the electric field is polarized along the static magnetic field. Conversely, a transition between two states with the same \hat{l}_z quantum number will produce light polarized along the magnetic field direction.

Selection rule on l

Next we consider the integral over θ . Here it is useful to consider properties of the angular parts of the wave function under coordinate inversion - in quantum mechanics we call this *Parity* inversion. In the end we are looking for the parity of the inside of the integral - if the parity is odd, then the integral will equal zero, otherwise we will get a value. So how might we find whether the wave functions are odd or even? There's a really nice trick to this, which I found in a book by J. J. Binney (it also appears in Bransden and Joachain). We consider only the angular parts of the wave function associated with the maximal value of \hat{l}_z . From our consideration of the angular momentum eigenfunctions of hydrogen, we know that these are given by

$$Y_{l,l}(\theta, \phi) = K \sin^l \theta e^{il\phi} \quad (7.156)$$

where K is a constant. When we invert parity in spherical co-ordinates, we make the transformation $\phi \rightarrow \pi + \phi$, $\theta \rightarrow \pi - \theta$ (if you are worried about this, go and look at the definition of spherical co-ordinates at the beginning of our treatment of Hydrogen). Since $\sin(\pi - \theta) = \sin(\theta)$, we find that

$$\hat{P} Y_{l,l}(\theta, \phi) = K \sin^l \theta e^{il\pi} e^{il\phi} = (-1)^l Y_{l,l}(\theta, \phi) \quad (7.157)$$

where \hat{P} is the parity operator. We see that for this type of state, the parity eigenvalue is just $(-1)^l$. We next note that all the angular momentum operators \hat{l}_x , \hat{l}_y , \hat{l}_z are even, and therefore \hat{l}_+ and \hat{l}_- are even. Since we find all other $Y_{l,m}(\theta, \phi)$ states by repeated application

of these operators, all other states with the same l will have the same parity. Thus we know that for any state with angular momentum eigenvalue l , the parity is $(-1)^l$.

Now let us return to our selection rule. $\cos(\pi - \theta) = -\cos(\theta)$, and therefore the product of parities for the two wave functions must be odd for the integral to be non-zero.¹ We therefore require that

$$(-1)^l (-1)^{l'} = (-1)^1 \quad (7.158)$$

and therefore

$$l = l' \pm 1 . \quad (7.159)$$

This is our second *selection* rule. The total electron angular momentum quantum number must change by one unit.²

Selection rule on n

The remaining cases are states which satisfy both of the previous rules. Are there any of these for which a transition does not happen? For this, we have to examine the integral over r . The integral is taken from 0 to ∞ , and r doesn't have a parity rule, since there are no negative values. We do know that if we didn't have the extra factor of r from the light, we would have orthogonality for different values of n . However, since we multiplied by this factor, we might expect that all the parts which contrived to cancel each other in the original integral don't in our current integral. We would be right. To check this argument, I looked at a few cases, and found none for which the integral is zero. We conclude that any change in principal quantum number n is possible.

Let me add a final note on selection rules for linear polarized light with polarization aligned with the magnetic field. It will only be emitted perpendicular to the magnetic field direction. This is because the electric field vector must be perpendicular to the magnetic field vector. This is an important factor in analyzing light emitted from atoms.

7.5.2 Circular polarization

At the beginning of the course, we considered different polarizations of light. We discussed linear polarization, but we also looked at circular polarization. The reason was because of its significance regarding the interaction of light with matter. Circular polarization is a sum of two linear polarization components, $\pi/2$ out of phase with each other. We can write it as

$$\frac{E_0}{\sqrt{2}} (\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}) \quad (7.160)$$

and thus

$$\mathbf{E}(0) \cdot \mathbf{r} = E_0(x \pm iy) = E_0 r (\cos(\phi) \sin(\theta) \pm i \sin(\phi) \sin(\theta)) \quad (7.161)$$

$$= E_0 r e^{\pm i\phi} \sin(\theta) \quad (7.162)$$

¹If you are thinking “surely cos is an even function?” then you would be right, but you would be talking about a different symmetry to the parity inversion. This is something it is worth being aware of - you can easily trip up.

²Our treatment so far actually only shows, that l and l' differ by an odd integer. More tough calculations then show, that this is really one. If you are into it, eg. have a look at F. Schwabl, Quantum Mechanics, p. 303ff.

Selection rule on m and l

On comparing this expression with example considered in the previous section, we quickly see that now the condition on the \hat{l}_z eigenvalues is

$$m' - m = \mp 1 \quad (7.163)$$

and since the operator again has odd parity with respect to inversion

$$l - l' = \pm 1 \quad (7.164)$$

Thus circularly polarized light is able to add or subtract one unit of angular momentum about z . The light must be carrying angular momentum, and it must be around the z axis! This is indeed the case. Circularly polarized light, with the electric field rotating around the z axis, does indeed carry one unit of angular momentum about this axis. In fact, photons always carries one unit of angular momentum; this is why $l - l'$ always has magnitude 1 above.

Chapter 8

Spin

Thus far in the course, every physical property which we have looked at, such as position, momentum, energy etc. have roles in both classical and quantum physics. We now come to a topic which was only discovered with the advent of quantum physics, and has no analogy in classical physics.

We call this property “intrinsic angular momentum” or “spin”, but it can not be explained (at least up until now) as anything physical spinning through space. The name comes because many of its features are shared with systems which do exhibit angular momentum. For instance, the spin gives rise to a magnetic dipole moment which interacts with magnetic fields in the same manner as the Zeeman effect which we considered in the previous section for a charged particle with orbital angular momentum. So here at last you should let go of all notions of the classical world, and indeed the maths we will use in representing spin now has all the features of more advanced quantum mechanics courses, with matrices used to represent operators rather than explicit mathematical forms acting through co-ordinates.

8.1 Experimental observations

8.1.1 Doublets in spectroscopy

What were the experimental observations which led physicists to arrive at this new physical property? The primary observations were more precise measurements of the spectral lines of atoms. In the hydrogen atom, the $n = 3 \rightarrow n = 2$ line of the Balmer series which has a wavelength of $\lambda = 656.3$ nm is seen to be split into two lines with a wavelength difference of $\Delta\lambda = 0.014$ nm. The puzzle is that the splitting is into two lines - a level with angular momentum l should always be split into an odd number ($2l + 1$) of states, and requires a magnetic field - what is observed in experiments is a splitting into two different energies, and exists without an external magnetic field being present. Similar effects are seen for a variety of atoms, including sodium (which is predicted to have very similar structure to hydrogen, as we will find out later in the course). For heavier atoms this splitting is larger.

8.1.2 Stern-Gerlach experiment

In 1922 O. Stern and W. Gerlach conducted experiments with beams of neutral silver atoms. A schematic of their experimental setup is shown in figure 8.1. They first created a beam

of silver atoms by heating an oven with silver inside. At high enough temperatures atoms effuse from the oven and are subsequently collimated by a sequence of apertures. This atomic beam is sent through an area with a strong *inhomogeneous* magnetic field where the major component of both the magnetic field and its gradient lie along the z direction ($B_z \gg B_x, B_y$ and $|\partial\mathbf{B}/\partial z| \gg |\partial\mathbf{B}/\partial x|, |\partial\mathbf{B}/\partial y|$) and then collected on a sheet of glass.

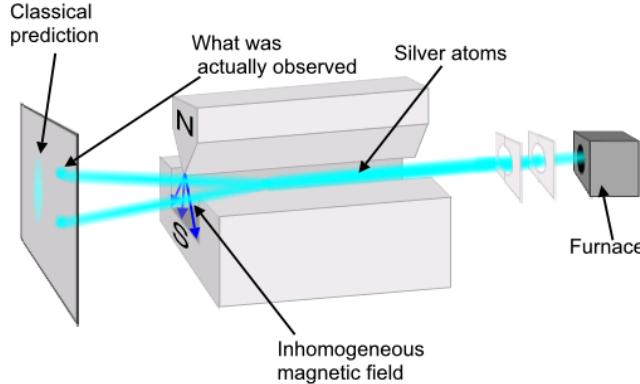


Figure 8.1: A schematic of the Stern-Gerlach experiment. Silver atoms effuse from an oven and are collimated before being passed through an inhomogeneous magnetic field. The final position of the atoms indicates the forces which they have felt in the field. Stern and Gerlach observed that the atoms split into two separate paths. This is not only different to what would be expected classically (where quantization of angular momentum does not occur and thus a continuous range of positions would be observed), but also different to our expectation from the quantum mechanics we considered thus far (where we would expect to see 1, 3, 5 etc. beams, and not 2).

The force on an atom with dipole moment μ in a magnetic field $\mathbf{B}(x, y, z)$ is given by the gradient of the energy

$$\mathbf{F} = -\nabla(-\mu \cdot \mathbf{B}) . \quad (8.1)$$

In the Stern-Gerlach experiment, with primary component of the magnetic field gradient in the z (vertical) direction, the force is

$$F\hat{\mathbf{z}} = \mu_z \frac{\partial \mathbf{B}}{\partial z} \hat{\mathbf{z}} \quad (8.2)$$

where μ_z is the z component of the magnetic dipole moment .

With no magnetic field the atomic beam is not deflected, and creates a silver dot on a “detection” sheet of glass opposite the source. The shape of this dot corresponds to the geometry of the collimating apertures.

When the inhomogeneous magnetic field is switched on the atoms split up into two separate beams, and appear at two different points on the glass. This is true for the silver atoms used by Stern and Gerlach, and is also true for hydrogen atoms in their ground state (as shown by T.E. Phipps and J.B. Taylor in 1927). The puzzle is again the factor 2. For atoms in a magnetic field with angular momentum l , we would expect to see $2l + 1$ beams, which can only yield an odd number. Instead we get an even number.

An additional piece of the puzzle can be derived from the magnetic moment measured in these experiments. For a range of different atoms, the magnetic moment was found to have the same value

$$\mu = \frac{e\hbar}{2m_e} \quad (8.3)$$

where m_e is the electron mass. This indicates that there might be some fundamental property which is common to a given particle (possibly the electron) rather than the structure of any particular atom.

8.1.3 Hypothesis of the electron spin

One explanation for the new quantum number was hypothesized by S. A. Goudsmit and G. E. Uhlenbeck in 1925:

- The electron behaves as if it had an intrinsic angular momentum, which has a z component which may take two discrete values. This intrinsic angular momentum is called *spin* and is written as an operator as $\hat{\mathbf{s}}$.

Since the degeneracy of a level with angular momentum l is $2l + 1$, the degeneracy of 2 seen in the Stern-Gerlach experiment spin suggests that the spin angular momentum should satisfy

$$(2s + 1) = 2 \quad (8.4)$$

and thus the total spin quantum number should be $s = 1/2$. By analogy with the real angular momentum case, the projection of the spin angular momentum onto the z axis can then take the values $\pm\hbar/2$.

Why can the electron spin not be associated with the electron physically spinning about its center of mass? Couldn't it be that the electron is finite in size? Let us examine this argument. The electron radius measured in X-ray scattering experiments is of order $r_e < 10^{-16}$ m. If all the charge of the electron were then orbiting at this radius, the angular momentum would be

$$I\omega = (m_e r^2) \frac{v}{r} = m_e r c \frac{v}{c} \quad (8.5)$$

where m_e is the electron mass, I is the moment of inertia and ω is the orbital angular frequency. If the angular momentum is to equal $\hbar/2$ we require

$$\frac{v}{c} = \frac{\hbar}{2m_e cr} \simeq 2000 \text{ (!!)} \quad (8.6)$$

which clearly violates relativity. Thus we reiterate that there exists no classical explanation for the phenomenon of the electron spin.

8.2 The quantum rules for spin

The characteristics of the electron spin \mathbf{s} correspond to those of the angular momentum \mathbf{l} . Therefore it is worth examining what features of the angular momentum operators give rise to the features we wish to replicate. Firstly we note that spin is a vector quantity, that is we write it as

$$\hat{\mathbf{s}} = (\hat{s}_x, \hat{s}_y, \hat{s}_z) . \quad (8.7)$$

From the argument of the previous section, we require that the eigenvalues of \hat{s}_z are $\pm\hbar/2$, with associated \hat{s}_z quantum numbers $m_s = \pm 1/2$. The quantum number used to represent

the angular momentum squared is $s = 1/2$, and thus the eigenvalue equations for \hat{s}_z and \hat{s}^2 (by analogy with ordinary angular momentum) are

$$\hat{s}_z \chi_{s,m_s} = \hbar m_s \chi_{s,m_s} \quad (8.8)$$

$$\hat{s}^2 \chi_{s,m_s} = \hbar^2 s(s+1) \chi_{s,m_s} \quad (8.9)$$

where $s = 1/2$ and $m_s = \pm 1/2$. There should be two eigenstates corresponding to the two different values of m_s . Since s is fixed at $1/2$ for the electron we shall drop this index in writing the wavefunctions χ in future. Since we claim that spin is not a property of our classical notions of space, we must be more abstract in representing the eigenfunctions of the spin operator. We write the spin eigenstates in vector form

$$\chi_{1/2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi_{-1/2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (8.10)$$

The operator for \hat{s}_z can then be written as a matrix

$$\hat{s}_z = \frac{\hbar}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (8.11)$$

which it is easy to check has eigenvalues $\hbar/2$ and $-\hbar/2$ and corresponding eigenvectors which are those given above. We can also see that

$$\hat{s}_z^2 = \frac{\hbar^2}{4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (8.12)$$

If we are to complete the analogy to orbital angular momentum, the operators for spin \hat{s}_x , \hat{s}_y , \hat{s}_z should have the same commutation relations to those for \hat{l}_x , \hat{l}_y and \hat{l}_z . This requires

$$[\hat{s}_x, \hat{s}_y] = i\hbar \hat{s}_z \quad (8.13)$$

$$[\hat{s}_y, \hat{s}_z] = i\hbar \hat{s}_x \quad (8.14)$$

$$[\hat{s}_x, \hat{s}_z] = -i\hbar \hat{s}_y. \quad (8.15)$$

One set of operators which satisfy this set of relations are

$$\hat{s}_x = \frac{\hbar}{2} \sigma_x, \quad \hat{s}_y = \frac{\hbar}{2} \sigma_y, \quad \hat{s}_z = \frac{\hbar}{2} \sigma_z \quad (8.16)$$

where σ_x , σ_y and σ_z are the *Pauli matrices*

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (8.17)$$

It is a matrix multiplication exercise to ensure that the commutation relations are satisfied by these operators. It is also worth noting that

$$\hat{s}_x^2 = \hat{s}_y^2 = \hat{s}_z^2 = \frac{\hbar^2}{4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (8.18)$$

and thus that

$$\hat{s}^2 = \hat{s}_x^2 + \hat{s}_y^2 + \hat{s}_z^2 = 3 \frac{\hbar^2}{4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (8.19)$$

which confirms that equation 8.9 is satisfied for both eigenstates.

Since the spin operators obey the same commutation relations as standard angular momentum, it follows that we can define creation and annihilation operators for spin. These are

$$\begin{aligned} \hat{s}_+ &= \hat{s}_x + i\hat{s}_y \\ \hat{s}_- &= \hat{s}_x - i\hat{s}_y. \end{aligned} \quad (8.20)$$

From the definition of the Pauli matrices, we can find the matrix form for these expressions

$$\begin{aligned}\hat{s}_+ &= \frac{\hbar}{2}(\sigma_x + i\sigma_y) = \frac{\hbar}{2} \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + i \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \right) \\ &= \frac{\hbar}{2} \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} = \hbar \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.\end{aligned}\quad (8.21)$$

This operator acting on $\chi_{-1/2}$ gives $\chi_{1/2}$ plus an additional factor of \hbar which disappears on normalization.

8.2.1 Matrix mechanics

At this point I should make a comment on the use of matrices in quantum mechanics. When we write the wavefunction as a sum over discrete eigenfunctions

$$\psi = \sum_i c_i \psi_i \quad (8.22)$$

the important information is contained in the set of coefficients c_i . So I could equally well write this state as a vector containing all these values. If I use this form, I have to figure out how operators act on the state. It turns out that each operator can be written in matrix form, and these matrices are Hermitian matrices for any physical observable. The matrices can be calculated from the given eigenstates using

$$O_{mn} = \int \psi_m^* \hat{O} \psi_n dx . \quad (8.23)$$

To act an operator written in matrix form on a given wavefunction, we multiply the matrix by the vector of coefficients. It is these methods which I have tried to avoid in teaching you this course, because I think these will be the focus of your more theoretical course next year. However in teaching spin they are unavoidable: Since we never write down a spatial function related to spin (there is none) we are forced to leave the states as vectors and use matrices for the operators.

8.2.2 Combining spin with the rest of physics

How do we combine the vector form of the spin eigenfunctions with the spatial wavefunctions which we already found for the hydrogen atom? Since spin is an intrinsic property of the electron, it is unrelated to the spatial wavefunction. Therefore for each spin eigenfunction we write a different spatial wavefunction. Thus the total wavefunction of the electron looks like

$$\psi(x) = \begin{pmatrix} \psi_{m_s=1/2}(x) \\ \psi_{m_s=-1/2}(x) \end{pmatrix} = \psi_{m_s=1/2}(x)\chi_{1/2} + \psi_{m_s=-1/2}(x)\chi_{-1/2} \quad (8.24)$$

using $\chi_{\pm 1/2}$ as defined in equation 8.10. If the electron is in a harmonic oscillator potential or a square well, then the spatial wavefunctions associated with the two spin components will be the same. However in the presence of *magnetic field gradients* the spatial wavefunctions will differ, due to the *magnetic moment of the electron which is associated with the electron spin*. These lead to forces on the electron which depend on the spin state. It is these forces which are responsible for the discrete spatial splitting observed in the Stern-Gerlach experiment.

Difference between spin and ordinary angular momentum

In the previous sections, we have consistently made analogies between spin and ordinary angular momentum. What is the difference between the two? The primary difference is that the spin angular momentum always has the same magnitude for the electron. The quantum number associated with this is $s = 1/2$. By contrast, for ordinary angular momentum there is no limit to the size - the quantum number l can be as large as we like. This is the essence of why the spin is only seen in quantum mechanics - there is no possibility that the spin of the electron becomes macroscopic, because it is always $\hbar/2$.

8.3 The intrinsic magnetic moment

From an experimental perspective, the arguments which led to our introducing spin were that atoms are observed to have magnetic moments which could not be predicted by the theory which we had previously developed. It is the magnetic moment which is the key feature of the spin of the electron in terms of its coupling to the outside world. The experimental results found that the magnetic moment has magnitude along any one direction

$$\mu = \frac{e\hbar}{2m_e} = \mu_B \quad (8.25)$$

where μ_B is called the Bohr magneton. Let us compare this value to that which we found in an earlier lecture for the magnetic moment due to orbital angular momentum. There we found

$$\hat{\boldsymbol{\mu}}_l = -\frac{e}{2m_e}\hat{\boldsymbol{l}} \quad (8.26)$$

where $\hat{\boldsymbol{l}} = (\hat{l}_x, \hat{l}_y, \hat{l}_z)$ and I have now added the subscript l to indicate that this is an expression for ordinary angular momentum. Remember that for a state with angular momentum $\hbar m$ about the z -axis, the component of the magnetic moment is

$$\mu_{l,z}(m) = -\frac{e\hbar}{2m_e}m = -\mu_B m \quad . \quad (8.27)$$

Now consider the magnetic moment of the spin angular momentum. If we were to use a direct analogy to the ordinary case, we would write

$$\mu_{s,z}(m_s) = -\mu_B m_s \quad . \quad (8.28)$$

Since $m_s = \pm 1/2$, we would find that we underestimated the magnetic moment by a factor of $1/2$. So instead we introduce the electron g -factor, and write

$$\mu_{s,z}(m_s) = -g_s\mu_B m_s \quad (8.29)$$

with $g_s \simeq 2$. It therefore makes sense to write the magnetic moment operator for the spin as

$$\hat{\boldsymbol{\mu}}_s = -\frac{g_s e}{2m_e}\hat{\boldsymbol{s}} \quad . \quad (8.30)$$

You may have noted the approximately equals sign in $g_s \simeq 2$. It is there because g_s is not 2 ! The measured value is $g_s = 2.0023193043622 \pm 0.00000000000015$ (the latest measurement was performed in the Gabrielse group at Harvard using a single electron trapped in a Penning trap like the one which you considered in the problem sets). This is one of the most precisely measured values in physics.

The difference between g_s and 2 is due to quantum electrodynamics, which is far beyond the scope of this course. Precision measurements of this difference are among the most stringent tests of quantum electrodynamics, providing a benchmark for theoretical calculations. It is possible to use the measurement of $g_s - 2$ for the electron to estimate the fine-structure constant α , and actually this is one of the best ways to predict this. Since α is not measured directly, theory is needed to connect $g_s - 2$ to α . The level of the experiments is such that the theory for this has to be extremely accurate and becomes very hard. The errors in the theoretical results are thus comparable if not larger than those of the experiment!

It is very important to remember to use the $g_s \simeq 2$ in problems related to spin, otherwise your answers will be wrong by a factor of 2!

8.3.1 The spin of other particles

The electron is not the only elementary particle with spin. The proton and neutron both also have spin 1/2. The corresponding magnetic moments μ_p and μ_n are both on the order of the nuclear magneton

$$\mu_K = \frac{e\hbar}{2m_p} \quad (8.31)$$

which differs from the Bohr magneton by the mass ratio $m_e/m_p = 1/1836$ with m_p the mass of the proton. Note that the neutron has a spin magnetic moment even though it does not have any electrical charge. The differences between μ_p and μ_n are covered by proton and neutron g -factors

$$g_p \sim -5.58\mu_K \quad (8.32)$$

$$g_n \simeq 3.826\mu_K \quad (8.33)$$

The spin of compound objects such as nuclei are made up of the sum of spins of their constituents. Therefore nuclei also have spin, and depending on the number of protons and neutrons the spin can take half-integer and integer values. All nuclei with spin have a corresponding magnetic moment in the order of μ_K . Because the magnetic moment of nuclei is 1/1836 of the magnetic moment for the electron, we shall leave nuclear spin to be covered in more advanced courses.

8.4 The spin-orbit interaction

Now let us look to see if we can recover the splitting of states that were observed for the hydrogen atom. The splitting comes about because the spin magnetic moment interacts with the magnetic field associated with the electric field of the nucleus. Why does this magnetic field exist? Here we have to remember relativity: if we move into a frame moving at velocity \mathbf{v} relative to a frame in which there exists a static electric field \mathbf{E} , we will experience a magnetic field given by

$$\mathbf{B} = -\frac{\mathbf{v} \times \mathbf{E}}{c^2} . \quad (8.34)$$

If the electron has angular momentum, then classically it is moving through the electric field due to the nucleus. This leads it to experience a magnetic field.

The relativistic physics of the spin-orbit interaction is complex, and not worth going into here. It is complex because the electron is accelerating, and thus we are not in inertial frames of reference. The best way to treat the energy levels of hydrogen and include relativity is to solve the relativistic counterpart to the Schrödinger equation, which was derived by Dirac and is thus called the Dirac equation. The solutions to this equation naturally produce spin, without any need to introduce it as an extra effect.

Instead of taking the full relativistic approach, we will use the Biot-Savart law to derive the relevant equations. You'll see that even then we need a fudge factor, to account for some effects in relativity. As a student, this made me deeply uncomfortable, so I went looking for the "real" derivation. It turns out that it's a pretty hard problem, taking pages and pages of relativistic calculations (by which point I should have gone away and started again from the Dirac equation). I never got through the whole thing (real life became more of a priority), and I ended up accepting the fudge factor. It is enough for me to know that someone calculated this once, and I no longer feel a strong need to do it myself.

8.4.1 Perturbation Hamiltonian: semi-classical derivation

Instead of working in the rest frame of the nucleus, which we have until now, let us imagine that we are in the rest frame of the electron. Classically the nucleus then appears to orbit around the electron. We can view this as the electron sitting at the center of a current loop of radius r , with a current of magnitude

$$I = \frac{Ze|\mathbf{v}|}{2\pi r} \quad (8.35)$$

where Ze is the charge on the nucleus and \mathbf{v} is the speed of the electron around the nucleus. The factor Z is an integer number which allows us to account for nuclei with greater nuclear charge than that of Hydrogen. The Biot-Savart law which will give us the field at the electron can be written

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int \frac{Idl \times \mathbf{r}}{r^3} . \quad (8.36)$$

where \mathbf{r} in this expression is the vector from the point at which the field is evaluated to the centre of the current I flowing through an element of length dl . Since the direction of the current loop is in the direction of the velocity, we can write

$$Idl = \frac{Ze\mathbf{v}}{2\pi r} dl \quad (8.37)$$

where $\int dl = 2\pi r$. The position of the nucleus is given relative to the electron by $-\mathbf{r}$, where \mathbf{r} is the vector from the position of the nucleus to the position of the electron. The relevant

quantity in the Biot-Savart law is the former, therefore we find

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{e\mathbf{v} \times (-\mathbf{r})}{r^3} . \quad (8.38)$$

We now note that the angular momentum of the electron is $\mathbf{l} = \mathbf{r} \times \mathbf{p}$ and thus

$$\mathbf{v} \times -\mathbf{r} = \frac{\mathbf{r} \times \mathbf{p}}{m_e} = \frac{\mathbf{l}}{m_e} \quad (8.39)$$

which gives a value for the field at the electron

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{Ze\mathbf{l}}{m_e r^3} . \quad (8.40)$$

This is the field when we are in the rest frame of the electron. In practice, we work in the rest frame of the nucleus. In transforming back between frames we find that there is an extra factor $1/2$, which is called the *Thomas precession factor*. This is the factor which I tried to chase down as a student. What you need to know is that it exists, and that it is due to a relativistic transformation between reference frames.

In the rest frame of the atom, the magnetic field at the electron is therefore

$$\mathbf{B} = \frac{1}{2} \frac{\mu_0}{4\pi} \frac{Ze\mathbf{l}}{m_e r^3} . \quad (8.41)$$

As you can see, the magnitude of the field is proportional to the orbital angular momentum of the electron about the atomic nucleus, and it points in the same direction as the orbital angular momentum.

The magnetic moment due to the intrinsic spin angular momentum of the electron will interact with this magnetic field. The interaction energy is given classically by

$$U = -\boldsymbol{\mu}_s \cdot \mathbf{B} \quad (8.42)$$

where $\boldsymbol{\mu}_s$ is the spin magnetic moment vector. Moving to a quantum mechanical picture, we must replace the angular momentum, spin magnetic moment and the radial co-ordinate by their quantum mechanical operators. We thus obtain a Hamiltonian describing the interaction which looks like

$$\delta\hat{H} = \frac{g_s e}{2m_e} \frac{1}{2} \frac{\mu_0}{4\pi} \frac{Ze}{m_e} \frac{1}{r^3} \hat{s} \cdot \hat{l} \quad (8.43)$$

$$= \frac{g_s \mu_0 Z e^2}{16\pi m_e^2} \frac{1}{r^3} \hat{s} \cdot \hat{l} \quad (8.44)$$

$$= \frac{g_s \mu_0 Z e^2}{16\pi m_e^2} \frac{1}{r^3} (\hat{s}_x \hat{l}_x + \hat{s}_y \hat{l}_y + \hat{s}_z \hat{l}_z) . \quad (8.45)$$

8.4.2 The spin-orbit energy level shifts

Set of states for perturbation theory: mathematical consideration

Now we will use perturbation theory to evaluate the energy level shifts due to this Hamiltonian. Here we must be very careful in view of the degeneracy of our original set of Hydrogen basis states. Remember that we should choose a starting set of eigenstates which are described by operators which commute with both the unperturbed and the perturbing Hamiltonian. Including now the spin operators, the quantum numbers which we have used to describe the energy eigenstates of hydrogen prior to adding this perturbation are

$$n, l, s, m, m_s . \quad (8.46)$$

The only quantum numbers which label different states which are degenerate with each other are those related to angular momentum, namely l, s, m, m_s (states with different n are not degenerate). All the operators corresponding to these quantities clearly commute with our original Hamiltonian, otherwise we would not be using them to describe eigenstates of energy of the original Hamiltonian. But we now have to examine whether they commute with the perturbation $\delta\hat{H}$. The operator parts are

$$\frac{1}{r^3}(\hat{s}_x \hat{l}_x + \hat{s}_y \hat{l}_y + \hat{s}_z \hat{l}_z) . \quad (8.47)$$

The $1/r^3$ part of this operator will commute with all of the angular momentum operators, since none of the the angular momentum operators act on the radial co-ordinate r . Thus this part of the perturbation will not affect which set of states to choose. Also note that since the orbital angular momentum operators act on spatial co-ordinates, they commute with the spin angular momentum operators, which don't.

Let us next examine $\hat{l}^2 = \hat{l}_x^2 + \hat{l}_y^2 + \hat{l}_z^2$. We should remember that $[\hat{l}_x^2, \hat{l}_x] = 0$, and this should be the same for y and z . Consider

$$\begin{aligned} [\hat{l}^2, \hat{l}_x] &= [\hat{l}_x^2, \hat{l}_x] + [\hat{l}_y^2, \hat{l}_x] + [\hat{l}_z^2, \hat{l}_x] \\ &= [\hat{l}_y^2, \hat{l}_x] + [\hat{l}_z^2, \hat{l}_x] \\ &= \hat{l}_y[\hat{l}_y, \hat{l}_x] + [\hat{l}_y, \hat{l}_x]\hat{l}_y + \hat{l}_z[\hat{l}_z, \hat{l}_x] + [\hat{l}_z, \hat{l}_x]\hat{l}_z \\ &= \hat{l}_y(-i\hbar\hat{l}_z) + (-i\hbar\hat{l}_z)\hat{l}_y + \hat{l}_z(i\hbar\hat{l}_y) + (i\hbar\hat{l}_y)\hat{l}_z \\ &= 0 . \end{aligned} \quad (8.48)$$

Since \hat{l}^2 is symmetric with respect to x, y, z , we immediately know that it commutes with all terms of the perturbing Hamiltonian $\delta\hat{H}$. Since the form of $\delta\hat{H}$ is the same for angular and orbital angular momentum operators, the same will hold for \hat{s}^2 .

What about \hat{l}_z ? Here we find

$$\begin{aligned} [\hat{l}_z, \hat{s}_x \hat{l}_x + \hat{s}_y \hat{l}_y + \hat{s}_z \hat{l}_z] &= \hat{s}_x [\hat{l}_z, \hat{l}_x] + \hat{s}_y [\hat{l}_z, \hat{l}_y] \\ &= \hat{s}_x (i\hbar\hat{l}_y) + \hat{s}_y (-i\hbar\hat{l}_x) \\ &= i\hbar(\hat{s}_x \hat{l}_y - \hat{s}_y \hat{l}_x) \end{aligned} \quad (8.49)$$

It does not commute! Therefore we should not use the eigenstates of \hat{l}_z for our perturbation theory.

Finally we should consider \hat{s}_z . We find

$$\begin{aligned} [\hat{s}_z, \hat{s}_x \hat{l}_x + \hat{s}_y \hat{l}_y + \hat{s}_z \hat{l}_z] &= [\hat{s}_z, \hat{s}_x] \hat{l}_x + [\hat{s}_z, \hat{s}_y] \hat{l}_y \\ &= (i\hbar\hat{s}_y) \hat{l}_x + (-i\hbar\hat{s}_x) \hat{l}_y \\ &= i\hbar(\hat{s}_y \hat{l}_x - \hat{l}_y \hat{s}_x) \end{aligned} \quad (8.50)$$

which also does not commute. Again we note that we should not use eigenstates of \hat{s}_z for this perturbation.

What operators could we use to describe our states? Look carefully at equations 8.49 and 8.50. You should notice that the right hand sides are exactly the same, apart from a minus sign. This implies that the operator

$$\hat{j}_z \equiv \hat{l}_z + \hat{s}_z \quad (8.51)$$

will commute with the perturbation. This is an operator which adds the z -components of the spin and ordinary angular momentum together! It is therefore the z -component of the

operator for the *total angular momentum* of the system. For a state which is an eigenstate of \hat{l}_z and \hat{s}_z with eigenvalues $\hbar m$ and $\hbar m_s$ respectively, the eigenvalue of the \hat{j}_z operator is

$$\hat{j}_z \psi_{l,m} \chi_{s,m_s} = (\hat{l}_z + \hat{s}_z) \psi_{l,m} \chi_{s,m_s} = \hbar(m + m_s) \psi_{l,m} \chi_{s,m_s} . \quad (8.52)$$

Since m can take all values from $-l \rightarrow l$ and m_s can take values $-1/2$ and $1/2$ we find that the eigenvalues of \hat{j}_z which we call m_j , can take all values between $-l - 1/2$ and $l + 1/2$.

Set of states for perturbation theory - physical consideration

Let us consider physically what is going on here. The perturbing Hamiltonian tells us that the energy depends on the relative alignment of the spin and the orbital angular momentum. Thus this energy will change if the angle between these two vector quantities changes. This means that the spin produces a *torque* which acts on the orbital angular momentum, and vice versa. We would therefore expect that the direction of the two vectors changes with respect to time and therefore the *eigenstates of $\hat{l}_x, \hat{l}_y, \hat{l}_z$ and $\hat{s}_x, \hat{s}_y, \hat{s}_z$ will be time dependent*. The implication is that *these eigenstates are not energy eigenstates*. Thus we cannot use these for the perturbation theory.

What about the magnitude of the spin and orbital angular momenta? Clearly the size of the spin of the electron cannot change - it is fixed at $\sqrt{\hbar^2(s(s+1))}$ with $s = 1/2$. I cannot think of a convincing physical argument for the magnitude of orbital angular momentum - it might have something to do with conservative forces (such as a B-field) but I cannot spot the link, and no book I've seen attempts to make any physical argument. The commutation of the operators will therefore have to be enough!

By contrast to the spin and orbital components of the angular momentum, the total angular momentum cannot change in magnitude or direction, because there is no external torque acting on the system (there are no external forces acting on the system at all). This means that any operator of the total angular momentum will have a set of eigenstates which are stationary states of the whole Hamiltonian (including the perturbation). These are therefore good basis states to use for perturbation theory.

Total angular momentum eigenvalues

In addition to \hat{j}_z , we can also define total angular momentum operators for x and y

$$\hat{j}_x \equiv \hat{l}_x + \hat{s}_x \quad (8.53)$$

$$\hat{j}_y \equiv \hat{l}_y + \hat{s}_y \quad (8.54)$$

which can be put together into a vector operator for the total angular momentum

$$\hat{j} \equiv (\hat{j}_x, \hat{j}_y, \hat{j}_z) = \hat{l} + \hat{s} . \quad (8.55)$$

Note that because $\hat{j}_x, \hat{j}_y, \hat{j}_z$ are sums of the spin and orbital components, they satisfy similar commutation relations

$$[\hat{j}_x, \hat{j}_y] = i\hbar \hat{j}_z \quad (8.56)$$

$$[\hat{j}_y, \hat{j}_z] = i\hbar \hat{j}_x \quad (8.57)$$

$$[\hat{j}^2, \hat{j}_z] = 0 . \quad (8.58)$$

The result of this is that we can follow through the whole procedure that we used for the orbital angular momentum operators when we were studying the eigenfunctions of Hydrogen

originally. There, once we had the eigenvalues of \hat{l}_z we were able to find the corresponding eigenvalues of \hat{l}^2 using only the commutation relations (we defined \hat{l}_+ , \hat{l}_- etc.). It therefore follows by analogy that the eigenvalues of \hat{j}^2 are $\hbar^2 j(j+1)$. The corresponding eigenvalues of \hat{j}_z are then $\hbar m_j$ and take any integer values between $-\hbar j$ and $\hbar j$.

The next question to ask is what values of j are possible for orbital angular momentum quantum number l and spin angular momentum operator s ? The rule for doing this is rather simple, but seems to cause all sorts of problems the first time anyone comes across it. The idea is the following. The maximum value of the angular momentum pointing along the z direction is given by the maximum value of the quantum number m , which is l . The spin can either point in the same direction ($m_s = 1/2$) or the opposite direction ($m_s = -1/2$). If the spin is aligned with the orbital angular momentum, the total angular momentum along the z direction is then

$$m_j = m + 1/2 . \quad (8.59)$$

Since in this case both the spin and orbital angular momentum are pointing along the z axis, this is the maximum value that m_j can take, and the value of j is thus

$$j = l + s . \quad (8.60)$$

If the spin and orbital angular momentum are anti-aligned, the maximum value of m_j is

$$m_j = m - 1/2 . \quad (8.61)$$

and thus

$$j = l - s \quad (8.62)$$

Thus for the case of a spin 1/2 particle there are two possibilities for j , one where the spin is aligned and one where it is anti-aligned with the orbital angular momentum.

Evaluation of the energy shift

We are now in a position to evaluate the energy shift due to the perturbing Hamiltonian. The basis states we use are mutual eigenstates of \hat{H}_0 , \hat{j}^2 , \hat{j}_z , \hat{l}^2 , \hat{s}^2 with quantum numbers $n, j(j+1), m_j, l, s$ respectively. We note that we can write

$$\begin{aligned} \hat{j}^2 &= (\hat{l} + \hat{s}) \cdot (\hat{l} + \hat{s}) \\ &= \hat{l}^2 + \hat{s}^2 + 2\hat{l} \cdot \hat{s} \end{aligned} \quad (8.63)$$

which can be re-arranged to give

$$\hat{l} \cdot \hat{s} = \frac{1}{2} (\hat{j}^2 - \hat{l}^2 - \hat{s}^2) . \quad (8.64)$$

The eigenvalue of this operator acting on a state $\psi_{n,j,m_j,l,s}$ is therefore

$$\frac{\hbar^2}{2} (j(j+1) - l(l+1) - s(s+1)) . \quad (8.65)$$

Returning to the perturbation Hamiltonian, we also need to find the expectation value for $1/r^3$ in one of these eigenstates. This involves lots of tedious algebra, so we will not go into it here. The result is

$$\left\langle \frac{1}{r^3} \right\rangle = \frac{1}{a_0^3} \frac{Z^3}{n^3 l(l+1/2)(l+1)} . \quad (8.66)$$

Putting all of these factors together, we find that the energy shift is

$$\delta E_{n,l,j,s} = \frac{g_s \mu_0 Z e^2}{16\pi m_e^2} \frac{Z^3}{a_0^3 n^3 l(l+1/2)(l+1)} \frac{\hbar^2}{2} (j(j+1) - l(l+1) - s(s+1)) \quad (8.67)$$

$$= \frac{g_s \mu_0 e^2 \hbar^2}{32\pi m_e^2 a_0^3} Z^4 \frac{j(j+1) - l(l+1) - s(s+1)}{n^3 l(l+1/2)(l+1)} \quad (8.68)$$

which using the definitions $a_0 = \hbar/(m_e ac) = e^2/(4\pi\epsilon_0 \hbar c)$ and $c^2 = 1/(\mu_0 \epsilon_0)$ gives

$$\begin{aligned} \delta E_{n,l,j,s} &= \frac{g_s \alpha \hbar^3 m_e^3 \alpha^3 c^4}{8c^2 m_e^2 \hbar^3} Z^4 \frac{j(j+1) - l(l+1) - s(s+1)}{n^3 l(l+1/2)(l+1)} \\ &= \frac{1}{4} \alpha^4 m_e c^2 Z^4 \frac{j(j+1) - l(l+1) - s(s+1)}{n^3 l(l+1/2)(l+1)} . \end{aligned} \quad (8.69)$$

There are several things to notice about this expression.

- The energy levels now depend on j and l and not just on n , as we had previously.
- The scaling goes rapidly with nuclear charge Z^4
- The expression is proportional to $\alpha^4 m_e c^2$, whereas the original energy levels are proportional to $\alpha^2 m_e c^2$. Thus the order of magnitude of this shift is smaller by a factor $\alpha^2 = 5.3 \times 10^{-5}$. Our use of perturbation theory is therefore valid.

The fact that the electron experiences a magnetic field is a relativistic effect. It is also the case that solutions to the relativistic equation for time evolution in quantum mechanics (which is called the Dirac equation) naturally include spin. Thus spin can be viewed as a relativistic effect.

8.5 Hydrogen fine structure

The spin-orbit interaction is not the only factor which affects the energy levels of hydrogen at these levels. We made several approximations in writing down the original Hamiltonian for the hydrogen atom. These can be accounted for in perturbation theory, and some are listed below.

- We assumed that the proton has infinite mass. Relaxing this assumption is simple - the mass of the electron in the above expressions should be replaced with the reduced mass of the proton and the electron. This is $\mu = m_e m_p / (m_e + m_p) \simeq m_e (1 - m_e/m_p)$. This multiplies the energies of all levels by the same factor, and the frequencies of transitions are also modified.
- We assumed that the electron was completely non-relativistic. We have already considered one of the effects which is neglected due to this assumption. We should also note that a relativistic expression for the energy is

$$\begin{aligned} E &= (p^2 c^2 + m^2 c^4)^{1/2} \\ &= mc^2(1 + p^2 c^2 / (2m^2 c^4) - p^4 c^4 / (8m^4 c^8) + \dots) \\ &= mc^2 + \frac{p^2}{2m} - \frac{p^4}{8m^3 c^2}, \end{aligned} \quad (8.70)$$

the third term of which can also be evaluated as a perturbation Hamiltonian using first-order perturbation theory.

Two other terms arise which have nothing to do with physics which we considered or did not consider so far. The first is called the Darwin term, and relates to rapid oscillations of the electron (called Zitterbewegung) which are beyond the level of our course. This term is taken into account by solutions of the Dirac equation, which naturally produce Zitterbewegung. It primarily shifts the energy levels of the $l = 0$ states. The second effect is described in section 8.5.2.

8.5.1 Combined results.

Performing perturbation theory on all the terms discussed above or solving the Dirac equation directly produces the same result for the energy levels of the hydrogen atom. The energy levels are given by

$$E_n^{\text{FS}} = -E_n \frac{Z^2 \alpha^2}{n} \left(\frac{1}{j+1/2} - \frac{3}{4n} \right) \quad (8.71)$$

where E_n are the energy levels obtained from our treatment of the gross structure. It is striking that l does not appear in this equation - look at the spin orbit term, which has l in many places. This means that the effects of the different l from each of the perturbations cancel each other out - the remaining dependence is on the total angular momentum represented by the quantum number j . Thus we note that the energy levels in hydrogen are governed by two quantum numbers, n and j . This indicates that there is much degeneracy in the hydrogen atom - indeed this is the case.

8.5.2 The Lamb shift

Finally, there is an effect called the Lamb shift which requires physics beyond even relativistic quantum mechanics to explain it. This was experimentally observed by Lamb and Rutherford

to shift the energy of the $l = 0$ eigenstates. This shift was eventually explained by the introduction of quantum electrodynamics. This shows the power of precision measurement for the study of fundamental science. By making an even more precise measurement, Lamb and Rutherford drove the production of new theories for explaining the universe.

Chapter 9

Atomic and molecular physics

We have now dealt in some depth with the structure of Hydrogen, the simplest of all atoms. It is a relatively simple problem because it involves two particles, an electron and a proton. Here we will see how the approach to the Hydrogen atom must be extended when atoms consist of nuclei with multiple electrons in orbitals. We will primarily work in the limit that the nucleus is a point charge with mass M_N , since the nuclear size is typically a million times smaller than the mean distance between the electrons and nucleus. The general problem of finding energy eigenstates and eigenvalues in the presence of many electrons is extremely hard, so I will restrict our aim to trying to provide a qualitative discussion of the relative energy levels and states of the atom rather than trying to produce exact solutions, as we did for Hydrogen. The ideas which we build up for Helium will then be applied to other many-electron atoms to give you some idea of how the periodic table arises from electronic structure.

9.1 Helium

The Helium is the simplest atom after Hydrogen, and the physics which arises in treating Helium gives us an idea of the basic considerations for atoms with larger numbers of electrons. The Helium nucleus consists of 2 protons and 2 neutrons, it therefore has charge $2e$. The Helium atom's Hamiltonian looks like

$$\hat{H} = \sum_{i=1}^2 \frac{\hat{\mathbf{p}}_i^2}{2m} - \sum_{i=1}^2 \frac{2e^2}{4\pi\epsilon_0 r_i} + \frac{e^2}{4\pi\epsilon_0 r_{12}} \quad (9.1)$$

where the first term is the electron's kinetic energy, the second the interaction between each electron and the nucleus, and the third describes the repulsion between the two electrons. \mathbf{p}_i is the momentum of the i th electron, r_i is its distance from the nucleus and r_{12} the distance between the two electrons. The major difference between this Hamiltonian and the one we considered for Hydrogen is the presence of the electron-electron repulsion. This makes the problem complex enough that it is generally better to use approximate methods than to try for an exact solution. Since there is no magnetic field, we will ignore the spin for now, and come back to its effects later.

9.1.1 Exchange symmetry

Let us start by ignoring the electron-electron repulsion (this is a bad approximation, but nevertheless it helps us to reveal some physics). In this case, each electron is independent, and feels the full charge of the nucleus, $2e$. At this level of approximation the wavefunctions for each electron will thus be similar to those of the Hydrogen atom (the mean radius will be reduced due to the increased nuclear charge).

We must be very careful about writing wave functions of multiple electrons. The electrons are identical particles, therefore if we exchange the two, we must end up with a wave function which produces identical physics. I will call this the *exchange* operation, and use the operator \hat{X} to specify its action. The exchange operator therefore has the action

$$\hat{X}\psi(1, 2) = \psi(2, 1) \quad (9.2)$$

where the 1, 2 in the arguments of the wavefunction ψ label the two electrons. Acting twice with this operator should recover the original wave function

$$\hat{X}\hat{X}\psi(1, 2) = \hat{X}\psi(2, 1) = \psi(1, 2) \quad (9.3)$$

We can use this to find the eigenvalues and eigenfunctions of this operator. These are the functions which satisfy the eigenvalue equation

$$\hat{X}\psi_X(1, 2) = X\psi_X(1, 2) \quad (9.4)$$

and thus we find that

$$\hat{X}\psi_X(1, 2) = X^2\psi_X(1, 2) \quad (9.5)$$

which since we must regain our original wave function on performing the exchange twice means that the eigenvalue X must be either plus or minus one (note that this is similar to the argument we made for Parity inversion when considering the finite square well, only now we are using exchange symmetry, and there we were using spatial symmetry and inverting the co-ordinate system).

Whether this value is plus or minus 1 is very important in quantum mechanics, since it leads to two fundamentally different classes of particle

- For spin 1/2 particles, such as electrons and protons, the value is -1 . These are *Fermions*.
- Spin 1 particles are *Bosons*. They have wavefunctions which are eigenstates of the exchange operator with eigenvalue $+1$.

Back to Helium. Electrons have spin 1/2 and are thus *Fermions*. Let us consider that the co-ordinate of electron 1 is \mathbf{r}_1 and that of electron 2 is \mathbf{r}_2 . They are in states 1 and 2, which are written as $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$. The wave function of both together will be a product of the single particle wave functions, for which eigenstates of the *spatial* Hamiltonian are products of the two. Naively we might think that two products are possible, which are most simply written as

$$\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2), \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1) . \quad (9.6)$$

Any superposition of these two might also be candidate wavefunctions, except that they do not have definite exchange symmetry (this is a physical requirement).

Now we note that the physics does not change if we exchange the electrons $\mathbf{r}_1 \leftrightarrow \mathbf{r}_2$, that is

$$[\hat{X}, \hat{H}] = 0 \quad (9.7)$$

and therefore we should be able to find simultaneous eigenstates of both operators. Note that the Hamiltonian *must* be symmetric with respect to exchange of the two electrons, otherwise the physics would definitely change, and the electrons could not possibly be identical. Applying the exchange operation to the eigenstate $\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)$, we find

$$\hat{X}\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) = \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1) \quad (9.8)$$

so clearly this is not an eigenstate of this operator. However two linear combinations *are* simultaneous eigenstates of both energy and exchange symmetry, these are

$$\psi_{21}^\pm = \frac{1}{\sqrt{2}} (\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) \pm \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)) \quad (9.9)$$

It seems that only one of these is allowed, though, because the wave function of our electrons, which are Fermions, should be antisymmetric with respect to exchange, and only the minus state has this property. At this point we should remember that the two electrons also have spin. The exchange principle is valid for the *total* wave function, not just the spatial part. Therefore not only does the spatial part of the wave function have to have definite symmetry, the spin part does too. Consider the ways in which the spin of the electrons could pair up. Either both could be aligned

$$\chi_{\uparrow_1 \uparrow_2}, \chi_{\downarrow_1 \downarrow_2} \quad (9.10)$$

or they could be anti-aligned, where the possible states include

$$\chi_- = \frac{1}{\sqrt{2}} (\chi_{\uparrow_1 \downarrow_2} - \chi_{\downarrow_1, \uparrow_2}) \quad (9.11)$$

$$\chi_+ = \frac{1}{\sqrt{2}} (\chi_{\uparrow_1 \downarrow_2} + \chi_{\downarrow_1, \uparrow_2}) \quad (9.12)$$

where again we chose linear combinations such that the spin part of the wave function would have definite symmetry. Clearly the spin wave function can also be either symmetric or antisymmetric with respect to exchange, where

$$\chi_{\uparrow_1 \uparrow_2}, \chi_{\downarrow_1, \downarrow_2} \text{ and } \chi_+ \quad (9.13)$$

are symmetric and

$$\chi_- \quad (9.14)$$

is anti-symmetric. The total wave function must therefore either be of the form

$$\frac{1}{\sqrt{2}} [\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) + \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)] \chi_- \quad (9.15)$$

or of the form

$$\frac{1}{\sqrt{2}} [\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) - \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)] \chi_+ \quad (9.16)$$

$$\frac{1}{\sqrt{2}} [\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) - \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)] \chi_{\uparrow_1 \uparrow_2} \quad (9.17)$$

$$\frac{1}{\sqrt{2}} [\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2) - \psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_1)] \chi_{\downarrow_1, \downarrow_2} \quad (9.18)$$

Note that this means that there is only one state which is symmetric in the spatial part, but there are three states with antisymmetric spatial parts. You should also note that if the spin states are the same, then the spatial eigenstates cannot be the same. **The electrons can not be in exactly the same state.** This is the *Pauli exclusion principle*, and it applies to any pair of *identical Fermions*.

The consequence of the Pauli exclusion principle for the Helium atom is that there is only one state allowed at the lowest principle quantum number. This state has the spins in different states, ie. anti-aligned

$$\psi_1(\mathbf{r}_2)\psi_1(\mathbf{r}_2)\chi_- \quad (9.19)$$

On the other hand, at each excited state there are two options for the spatial symmetry, one where it is symmetric (with one eigenstate), and one where the spatial symmetry is antisymmetric (with three eigenstates).

The energies of the antisymmetric and symmetric spatial states will be different due to the electron-electron repulsion. As a first step we can make an estimate of which of these spatial wavefunctions has the higher energy. Note that as $\mathbf{r}_1 \rightarrow \mathbf{r}_2$ the two terms in the wavefunction become similar to each other and thus the antisymmetric wavefunction tends to zero. By contrast, the spatially symmetric wavefunction tends to $2\psi_1(\mathbf{r}_2)\psi_2(\mathbf{r}_2)$ as $\mathbf{r}_1 \rightarrow \mathbf{r}_2$. Thus the amplitude to find the electrons close to each other is much higher for the spatially symmetric wavefunction than for the anti-symmetric wavefunction. The result is that the electron-electron repulsion energy is higher for the symmetric spatial states. The result is that the state which is spatially symmetric states and has an antisymmetric spin wavefunction lies at a higher energy than the three spatially antisymmetric states with the symmetric spin states.

9.2 Quantum statistics

Since we are covering the properties of different types of particles with respect to their indistinguishability, it is worth returning to look at the implications for statistical physics. In particular, we now know that the rules for filling up states is different for bosons and for fermions. The case we looked at earlier in the course was for the “classical” case of indistinguishable particles. Here we were able to derive the number of possible microstates for n_i particles to get into a “bundle” of g_i states with energy ϵ_i . The result was

$$\Omega_i = \frac{(n_i + g_i - 1)!}{n_i!(g_i - 1)!} . \quad (9.20)$$

In the previous part of the course, we then made the approximation that no two particles are in the same state, which is a good approximation for $g_i \gg n_i$. Now let us see what happens when we don’t make this approximation, which is the realm of “quantum” statistical mechanics. Here the quantum refers to whether we have to worry if the particles are Fermions or Bosons. We only need to worry if two particles might try to go into the same state (which is not allowed for Fermions).

9.2.1 Bose-Einstein distribution

In this section we will allow two particles into the same state, which means that we are looking at the case of Bosons. Here equation 9.20 is valid, but we are not going to take the limit that $g_i \gg n_i$ as we did previously. We will find occupation numbers for the macrostate with the largest number of microstates with the constraints that the total energy and the total particle number are fixed at E and N respectively. Finding this macrostate involves maximizing

$$\ln \Omega = \ln \left(\prod_i \Omega_i \right) = \sum_i \ln \Omega_i \quad (9.21)$$

with respect to the occupations n_i . As we did previously, we introduce the Lagrange multipliers β and α and find the solution to

$$\frac{\partial}{\partial n_i} \left(\sum_i \ln(\Omega_i) - \beta \left(\sum_i \epsilon_i n_i - E \right) - \alpha \left(\sum_i n_i - N \right) \right) = 0 . \quad (9.22)$$

As in the cases we studied in the earlier part of the course, it helps to use the Stirling approximation, restricting ourselves to cases where $n_i, g_i \gg 1$ (note that this is a different approximation to the “classical” approximation $g_i \gg n_i$). This allows the expansion

$$\begin{aligned} \ln \Omega_i &\simeq (n_i + g_i - 1) \ln(n_i + g_i - 1) - (n_i + g_i - 1) \\ &\quad - n_i \ln n_i + n_i - (g_i - 1) \ln(g_i - 1) + (g_i - 1) \end{aligned} \quad (9.23)$$

which simplifies the differentiation with respect to the n_i , resulting in

$$\begin{aligned} 0 &= \ln(n_i + g_i - 1) + 1 - 1 - \ln(n_i) - 1 + 1 - \beta \epsilon_i - \alpha \\ &= \ln \frac{n_i + g_i - 1}{n_i} - \beta \epsilon_i - \alpha . \end{aligned} \quad (9.24)$$

This expression can be re-arranged to get an expression for the n_i

$$\begin{aligned} n_i &= \frac{g_i - 1}{e^{\beta \epsilon_i + \alpha} - 1} \\ &\simeq \frac{g_i}{e^{\beta \epsilon_i + \alpha} - 1} \end{aligned} \quad (9.25)$$

where in the last line I again used the approximation that $g_i \gg 1$. This expression is the renowned *Bose-Einstein* distribution. It is common to make the substitutions $\beta = 1/(k_B T)$, $\alpha = -\mu/(k_B T)$ where μ is the *chemical potential*. The chemical potential gives the increase in energy of a system when an extra particle is added to it at constant pressure and entropy. It is defined thermodynamically by

$$\mu = \left(\frac{\partial U}{\partial N} \right)_{V,S} . \quad (9.26)$$

Note that the chemical potential is related to the constraint that the particle number remains fixed. In the case of photons, the Bose-Einstein distribution agrees with the Planck distribution if $\mu = 0$. This implies that the number of photons is not conserved. This is indeed the case: thermal equilibrium requires the photons to be absorbed at one frequency and emitted at other frequencies; energy conservation then requires that the photon number would not be conserved.

An important consequence of Bose-Einstein statistics is that at a critical but non-zero temperature the particles all go into the lowest energy state available. This is called Bose-Einstein condensation, and was realized with neutral atoms in 1995 (earning W. Ketterle, C. Wieman and E. Cornell the 2001 Nobel Prize). The control and study of ultracold atoms close to the critical temperature at which condensation happens forms an important research field in modern physics.

9.2.2 Fermi-Dirac distribution

The case of Fermions is very different. We have to revise how to fit n_i particles into a bundle of g_i states while remembering that no two particles can go into the same state (the Pauli exclusion principle). The number of possible microstates is now given by the Binomial coefficients

$$\Omega_i = \frac{g_i!}{n_i!(g_i - n_i)!} \quad (9.27)$$

and thus the total number of states for a set of occupations $\{n_i\}$ of a set of bundles $\{g_i\}$ is

$$\Omega = \prod_i \frac{g_i!}{n_i!(g_i - n_i)!} . \quad (9.28)$$

From this starting point, we can again find the occupation numbers corresponding to the macrostate with the highest number of microstates. We use the same approach, again using the constraints that $E = \sum_i \epsilon_i n_i$ and $N = \sum_i n_i$. Including the Lagrange multipliers associated with these constraints, the maximization is

$$\frac{\partial}{\partial n_i} \left(\sum_i \ln(\Omega_i) - \beta \left(\sum_i \epsilon_i n_i - E \right) - \alpha \left(\sum_i n_i - N \right) \right) = 0 . \quad (9.29)$$

To simplify the logarithmic term, we make the approximation $n_i, g_i \gg 1$ and use the Stirling rule for the factorials, which on differentiating with respect to n_i , gives the set of equations

$$-\ln n_i - 1 + 1 + \ln(g_i - n_i) + 1 - 1 - \beta\epsilon_i - \alpha = 0 . \quad (9.30)$$

Re-arranging this expression gives us the *Fermi-Dirac* distribution

$$n_i = \frac{g_i}{e^{\beta\epsilon_i + \alpha} + 1} . \quad (9.31)$$

Here we can again make the substitutions $\beta = 1/(k_B T)$ and $\alpha = -\mu/(k_B T)$. Note that the only difference between this and the Bose-Einstein distribution which we considered earlier is the positive sign in the denominator. This has very important consequences though.

The first feature which is worth noting about the Fermi-Dirac distribution is the point at which $n_i/g_i = 1/2$. This is given by

$$\frac{1}{2} = \frac{1}{e^{\beta(\epsilon_i - \mu)} + 1} \quad (9.32)$$

which can be re-arranged to give

$$e^{\beta(\epsilon_i - \mu)} = 1 \quad , \quad (9.33)$$

which is satisfied for $\epsilon_F = \mu$. This value is known as the Fermi energy. The significance is that all levels lower than the Fermi energy are filled with probability greater than 1/2, and all levels with energies higher than the Fermi energy have a probability to be filled which is less than 1/2. At zero temperature, all states which have energies lower than the Fermi energy will be filled, while all states with energy above the Fermi energy are empty. The distribution is illustrated for a range of temperatures in figure 9.1. It can be seen that the temperature affects how rapidly the change between high and low occupancy takes place.

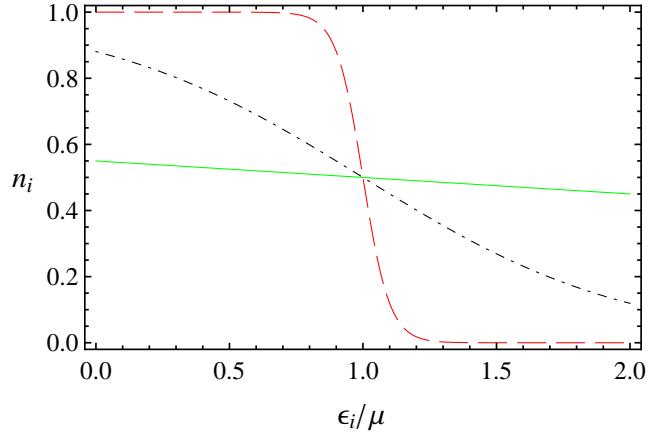


Figure 9.1: The occupation of states n_i/g_i according to the Fermi-Dirac distribution, given for $k_B T = 0.05\mu$ (red, dashed), $k_B T = 0.5\mu$ (black, dot-dashed) and $k_B T = 5\mu$ (green, solid). The energy of the states is given in units of the Fermi energy defined in the text. The transition between fully occupied and empty states is much sharper as the temperature is reduced.

9.3 Many-Electron atoms

The structure of multi-electron atoms is extremely complex, and the details of calculations are not only way beyond the level of this course, but many other courses too. This is a difficult many-body problem, which is solved by approximate numerical methods (which have nevertheless been used to make extremely accurate predictions). Here we will summarize the basic physics behind these approaches, which will allow us to qualitatively explain the form of the periodic table.

The Hamiltonian describing a many-electron atom, under the approximation that the nucleus is infinitely heavy, is given by

$$\hat{H} = \sum_{i=1}^N \left\{ \frac{\hat{\mathbf{p}}_i^2}{2m_e} - \frac{Ze^2}{4\pi\epsilon_0 r_i} \right\} + \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0 r_{ij}} \quad (9.34)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between electrons i and j , $\hat{\mathbf{p}}_i$ is the momentum operator for the i th electron and r_i is its distance from the nucleus. This is the exact Hamiltonian, which it would be very hard to solve.

9.3.1 The central field approximation - configurations

One procedure for simplifying this problem is to make the *central-field-approximation*. In this approach, the first approximation is to treat each electron independently, and to absorb all of the radial terms in the potential into a central field $U(r)$ which is the same for each electron. This allows us to separate out the exact Hamiltonian in the following way

$$\hat{H} = \hat{H}_c + \hat{H}_{nc} \quad (9.35)$$

where the two parts contain the central and the non-central parts of the Hamiltonian, given by

$$\hat{H}_c = \sum_i^N \left\{ \frac{\hat{\mathbf{p}}_i^2}{2m_e} + U(r_i) \right\} \quad (9.36)$$

and

$$\hat{H}_{nc} = \sum_i^N \left\{ \frac{Ze^2}{4\pi\epsilon_0 r_i} - U(r_i) \right\} + \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0 r_{ij}} \quad (9.37)$$

respectively. The basic idea is to solve for the eigenfunctions of \hat{H}_c first, and then to subsequently add \hat{H}_{nc} as a perturbation. This will only be possible for a particular form of $U(r_i)$. Note that since the potential term of \hat{H}_c is spherically symmetric, the angular momentum eigenfunctions in this approximation will be the same as those for Hydrogen, thus the i th electron will have an orbital angular momentum with eigenvalues labeled by the quantum numbers of total and z angular momentum, which we write as l_i and m_i respectively. We will retain our requirement that the electron is bound, and thus we again obtain a principle quantum number n_i for each electron. On top of this, each electron has spin, which we will denote by the spin quantum number $m_{s,i}$.

Think for a moment about the expected form for the central potential $U(r_i)$ from the perspective of one of the electrons. Very close to the nucleus, we would expect that the electron “sees” the full nuclear charge, and so the central potential is $-Ze^2/(4\pi\epsilon_0 r_i)$. On the other hand, if the electron is far from the nucleus, all the other electrons have high probability to be relatively close to the nucleus, and should therefore screen the charge

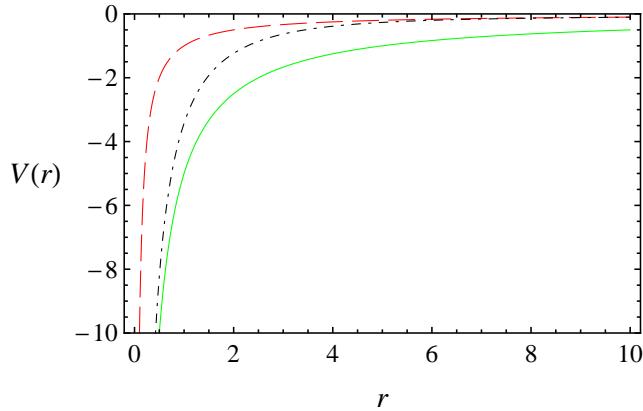


Figure 9.2: The form of the central potential falls between two extremes. At distances close to the nucleus, the electron will be exposed to the full nuclear charge, and thus experiences a potential proportional to $-Z/r$ (green, solid curve). For an electron far outside the mean radius of the other electrons, the charge of the nucleus is well screened, and the potential experienced is of the form $-1/r$ (red, dashed). The central potential should reproduce this behavior in these two limits, but will make a transition between the two at intermediate distances. I invented a form which does this to illustrate this point - this is represented in this figure by the dot-dashed, black curve.

of the nucleus. The net charge experienced by this outer electron will give an attractive potential $-e^2/(4\pi\epsilon_0 r_i)$. Between these two extremes, we should expect that the potential is something between these two cases. This is represented graphically in figure 9.2.

The Pauli exclusion principle means that no two electrons can be in the same state. If we imposed this requirement on the electrons in our atom, it means that no two electrons can share all of n_i, l_i, m_i and $m_{s,i}$. Since each electron can have two values of $m_{s,i}$, only two electrons can share their values of n_i, l_i and m_i . Now think back to the Hydrogen atom. The allowable values of l for a given n are $1 \rightarrow n - 1$. For each value of l we have $2l + 1$ allowable values of m . This leads to degeneracy of n^2 for each value of n in hydrogen.

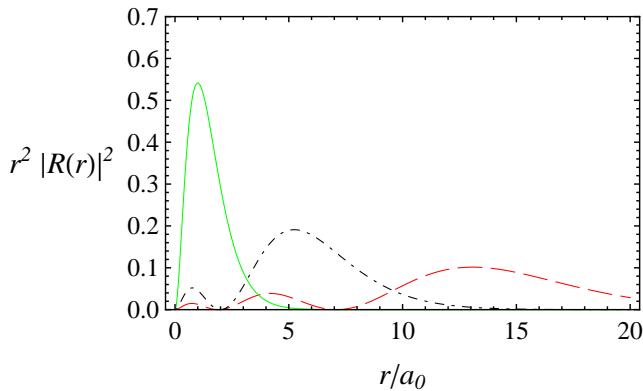


Figure 9.3: The probability of finding an electron at distance r/a_0 for fixed l and $n = 3, 2, 1$ (red dashed, black dot-dashed, green solid) for hydrogen. The difference in the expectation value of the radius is clear from these distributions.

For many-electron atoms the situation is more complicated, because the central field also contains spherically symmetric parts of the electron-electron repulsion. The value of n has the largest effect on the energy of an electron, and the states with higher n have mean radii which are much larger than those with lower n .

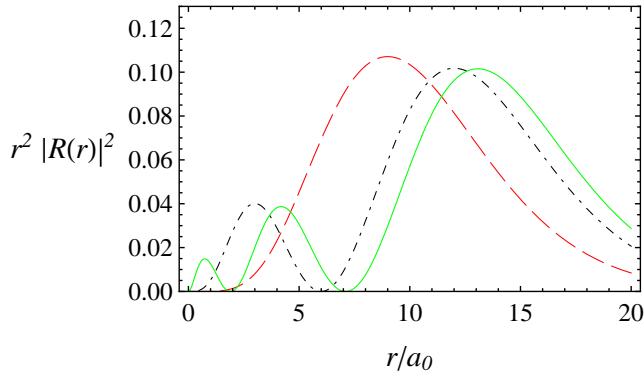


Figure 9.4: The probability of finding an electron at distance r/a_0 for fixed n and $l = 2, 1, 0$ (red dashed, black dot-dashed, green solid) for hydrogen. Though it is not terribly clear, the higher l states have a lower probability of being close to the nucleus. For many-electron atoms in the central field approximation this means that they experience less of the unscreened nuclear charge. The fact that this is much less obvious than for the case of fixed l and different n shown in figure 9.3 indicates that the difference in energies between states of different l will be much smaller than for different n .

For fixed n , the $l = 0$ state has the highest probability of being closest to the nucleus, and as l increases the mean radius increases (see figure 9.4 for an example). Therefore for the case of fixed n , states with higher values of l have higher energy. The lowest energy states for an electron will therefore be $n = 1, l = 0$, followed by $n = 2, l = 0$ and $n = 2, l = 1$. Two electrons can occupy the $l = 0$ states, so long as their spins differ. For $l = 1$, the degeneracy due to different m values is $2l + 1 = 3$, and combined with a degeneracy of two for spin, this leads to 6 electrons being allowed to occupy these states. In atomic notation, we call $l = 0$ an s orbital and $l = 1$ a p orbital, with $l = 2, 3, 4$ given labels of d, f, g respectively. At this level of approximation, the energy levels are well described by a list of the n and l values occupied by all the electrons, which we call the electron *configuration*. For Hydrogen, Helium, Lithium and Beryllium these lists look like

$$\begin{aligned} &1s \\ &1s^2 \\ &1s^2 2s \\ &1s^2 2s^2 \end{aligned}$$

where the superscript denotes the number of electrons with this set of quantum numbers. For heavier atoms the lists get longer because there are more electrons. The orbitals with other values of angular momentum also get filled. Examples for oxygen, aluminium and calcium are

$$\begin{aligned} &1s^2 2s^2 2p^4 \\ &1s^2 2s^2 2p^6 3s^2 3p \\ &1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 \end{aligned} \tag{9.38}$$

9.3.2 L-S coupling

At the next level of approximation we re-introduce the non-central terms in the potential. These come from residual terms in the electron-electron repulsion. Here we note that the electron orbitals have a well defined shape, with the orientation dependent on the value of m (which tells us the projection of the angular momentum along a specified direction). For

a given configuration, there are many different sets of relative orientations of the angular momenta of the different electrons relative to each other. Each of these sets will produce a different energy resulting from the non-spherically symmetric part of the interaction between the electrons. This means that the energies of the levels partially lose their degeneracy dependent on the sum of all the angular momentum vectors. That is for different total orbital angular momentum

$$\hat{L} = \sum_i \hat{l}_i , \quad (9.39)$$

the energies of the state will be different.

In addition to the orbital angular momentum, the symmetry properties of the wavefunctions for each configuration with respect to exchange of two electrons depends on their relative spin. Remember the case of the helium atom - the exchange symmetry which was required of the whole wavefunction imposed symmetry with respect to exchange on the spatial part. This in turn affected the mean separation of electrons and thus the mutual Coulomb repulsion.

The same physics is behind the difference in energy between states with different combinations of spin quantum numbers in many-electron atoms. Here we find that the energies of the levels have different energies dependent on the total spin vector

$$\hat{S} = \sum_i \hat{s}_i \quad (9.40)$$

for which we have a total angular momentum eigenvalue equation

$$\hat{S}^2 \psi = \hbar^2 S(S+1) \psi \quad (9.41)$$

with eigenvalue specified by the number S .

It is worth noting the way in which the electron states fill up with respect to angular momentum. The states with a single value of the orbital angular momentum quantum number are degenerate but all have different values of m . Therefore if all of these are filled the state of all the electrons together is spherically symmetric, and has net angular momentum zero. A similar argument holds for the spin quantum number. Therefore fully filled orbitals (an orbital contains all states with a given n and l) have zero angular momentum (both for spin and orbital angular momentum), examples being s^2 and p^6 . The total orbital angular momentum of all the electrons L is then given by the sum of the orbital angular momentum of the extra electrons which are outside the filled orbitals.

Since states with different L and S have different energies, it is useful to provide an extra labeling on top of the configuration. This is written as

$$(2S+1)L \quad (9.42)$$

where L is again represented by letters S, P, D, F corresponding to $L = 0, 1, 2, 3$ respectively. As examples, the three lowest energy levels of the calcium atom are given by

$$\begin{aligned} & 1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 \quad ^1S \\ & 1s^2 2s^2 2p^6 3s^2 3p^6 4s 4p \quad ^1P \\ & 1s^2 2s^2 2p^6 3s^2 3p^6 4s 4p \quad ^3P \end{aligned}$$

9.3.3 Spin-orbit coupling

I should mention that spin-orbit coupling also occurs in many-electron atoms and is the next level of approximation after L-S coupling. Here, each electron is travelling through the

electric field of the nucleus plus all of the other electrons. The spin magnetic moment of the electron will interact with this field, leading to a Hamiltonian which looks like

$$\hat{H}_{SO} = \sum_i \xi_i \hat{l}_i \cdot \hat{s}_i . \quad (9.43)$$

It is not obvious and require physics beyond this course to show that this Hamiltonian produces a perturbation of states consistent with

$$\hat{H}_{SO} = \xi \hat{\mathbf{L}} \cdot \hat{\mathbf{S}} . \quad (9.44)$$

In a similar manner to spin-orbit coupling in hydrogen, degenerate perturbation theory requires that the states used must be eigenstates of a physical observable which is a constant of the motion (ie. represented by operators which commute with the perturbation Hamiltonian). This leads to the introduction of the operator

$$\hat{\mathbf{J}} = \hat{\mathbf{L}} + \hat{\mathbf{S}} \quad (9.45)$$

which has an eigenvalue equation

$$\hat{J}^2 \psi_J = \hbar^2 J(J+1) \psi_J . \quad (9.46)$$

The energy levels with a single value of L, S are then split up according to a term proportional to

$$\delta E = \frac{A}{2} (J(J+1) - L(L+1) - S(S+1)) \quad (9.47)$$

in a similar manner to hydrogen. A is a constant which must be determined numerically. The associated spectroscopic notation is

$$(2S+1)L_J \quad (9.48)$$

where J is given as a number (the only number represented as a letter in this notation is for L).

9.3.4 $j-j$ coupling

It is worth pointing out that in heavy atoms, the spin-orbit coupling is larger in size than the residual electrostatic interaction. This means that the additions to the Hamiltonian should be performed in a different order, ie. with perturbation theory for spin-orbit coupling applied before that for the residual Coulomb repulsion. The choice of constants of the motion is thus different, and this leads to a different labeling system to that described above. You do not need to know the rules for this, but you should be aware that this can occur.

9.4 The Periodic Table

The periodic table is shown again (you saw it earlier in the lectures) in figure 9.5. You should now be able to look at it with a fresh viewpoint. We have just seen that two electrons go into s states, 6 into p etc.. Looking at the first two three rows of the periodic table, we see that there are 2 elements on the first row, which have electrons in the 1s state. On the next row, the 2s and 2p orbitals are gradually filled, so there are 8 different elements on this line. The next line seems to show the same behavior. For heavier atoms it is not so clear what the pattern might be, but we note that there are $2 + 6 + 10 = 18$ elements on each row, which corresponds to the summed number of states in the s , p and d orbitals.

Legend

- Atomic number
- Symbol
- Atomic weight
- Series
- Alkali metals
- Alkaline earth metal
- Inner transition metals
- Lanthanides
- Actinides
- Metals
- Transition metals
- Nonmetals
- Halogens
- Noble gases

Period

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	H																	He
2	Li	Be											B	C	N	O	F	Ne
3	Na	Mg											Al	Si	P	S	Cl	Ar
4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
6	Cs	Ba	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn	
7	Fr	Ra	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Uut	Uup	Uuh	Uus	Uuo		

Gruppe																	
3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36		
Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr		
38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
55	56	57-71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86
Cs	Ba	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn	
87	88	89-103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118
Fr	Ra	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Uut	Uup	Uuh	Uus	Uuo		

Lanthanides														
57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
138.91	140.12	140.91	144.24	146.90	150.36	151.96	157.25	158.93	162.50	164.93	167.26	168.93	173.05	174.97
Actinides														
Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr
(227)	232.04	231.04	238.03	237.05	(244.10)	(243.10)	(247.10)	(251.10)	(254.10)	(257.10)	(258)	(259)	(260)	

Figure 9.5: The periodic table.

9.4.1 The alkali group

The simplest of the elements to understand spectroscopically are the Alkali elements, which make up the left-most column of the periodic table. These share a common feature, which is that all electrons bar one are in filled orbitals with the additional electron in the state with a higher value of n than all of the others.

Since the filled orbitals spend more time closer to the nucleus than the extra electron, the state of the single electron completely determines the chemical properties of the atom. This electron sits outside a core composed of both the nucleus (with Z protons) and the other ($Z - 1$) electrons, therefore this core has a net charge of e .

In this sense the alkali atoms are similar to hydrogen. The spin of the outer electron is fixed at $S = 1/2$ and the ground state is $L = 0$. However contrary to the case of hydrogen, the states of different angular momentum are not degenerate. This is because the electron can penetrate the region inside the other electrons, where the potential is much deeper than that of a singly charged nucleus. The extent to which this electron penetrates the inner shells determines the energy level of the atom. Now you must think about the shape of the hydrogen wavefunctions. The $l = 0$ states had significant probability for the electron to be

found near the core, whereas the higher l states have an r^l dependence which ensures they are zero at the core. For this reason the $l = 0$ are attracted by the nuclear charge more strongly and thus have lower energy.

The alkali elements are highly chemically reactive. This is because the outer electron is well screened from the nucleus of the atom by the electrons in the lower n states. This electron is therefore relatively weakly bound, and can be “grabbed” by entering an available quantum state belonging to a nearby atom of another type. This is what is necessary for two atoms to form a molecule.

9.4.2 The inert gases

The other column of the periodic table which it is worth emphasizing is the right-hand column, which are called the inert gases. If these atoms are in their ground state, the highest energy electrons fill the outermost orbital. These orbitals will have a similar mean radius to all of the other electrons in the outer shell, hence the screening of the nuclear charge is poor. As a result, even the highest energy electrons are significantly exposed to the core charge of the nucleus. For a chemical reaction to occur, an additional atom must provide an energy state with a lower energy - since the highest energy in the inert gases are already at low energies (strongly bound) this is difficult to achieve. For this reason these gases are inert - it is very difficult to force them to undergo chemical reactions.

9.5 Adding angular momenta

Throughout the last few lectures, we have been adding different components of angular momentum. I don't want to teach you the formal theory for addition of angular momentum, because that is an advanced topic which you will cover in further courses. You should have some idea of what the rules are though, because these are important in forming possible values of S, L and J for a given atom from a set of quantum numbers for the individual electrons; l_i, s_i .

The way to think about this problem is the following. Let us imagine that you have two objects with angular momentum l_1 and l_2 . If we pick any one direction, we have a number of possible values of projections of these angular momentum on to that direction. These are $m_i = l_i, l_i - 1, \dots, -l_i$ where $i = 1, 2$ label the two objects. In order to add the two angular momenta together, first pick the longer of the two. I will choose this to be $l_1 > l_2$. Choose your axis system to line up with this vector. The projection of this vector onto the axis is therefore $m_1 = l_1$.

We now consider the projection of the other angular momentum onto this axis. The possible values for the projection are $m_2 = l_2, l_2 - 1, \dots, -l_2$. The length of the sum of the two angular momenta are then given by $l_1 + m_2$ for all these different values. Thus the possible values for the quantum number L for total angular momentum operator given by

$$\hat{L} = \hat{l}_1 + \hat{l}_2 \quad (9.49)$$

are

$$L = l_1 + m_2 \quad (9.50)$$

where $m_2 = l_2, l_2 - 1, \dots, -l_2$. Since the components of \hat{L} obey the same form of commutation rules as the components of both \hat{l}_i , for any given value of L the projection on any chosen axis is given by $M_L = L, L - 1, \dots, -L$.

Though I used l, L in the above example, the same method is used for adding any form of angular momentum, such as spin or total angular momentum, or indeed for adding different types of angular momentum. Some examples follow.

Example: Addition of angular momentum for $l_1 = 2, l_2 = 3$. The larger quantum number is l_2 , so let us choose an axis aligned with this vector. We therefore have $m_2 = 3$. The possible projections of angular momentum \hat{l}_1 onto this axis are then $m_1 = 2, 1, 0, -1, -2$. Thus the possible values of L are $5, 4, 3, 2, 1$.

Example: Addition of angular momentum for $s_1 = 1/2, s_2 = 1/2$. These are of equal magnitude, therefore we choose to align our axis with either. Let us choose s_1 , and thus $m_{s,1} = 1/2$. The second spin can have projections $m_{s,2} = \pm 1/2$ along this axis, so finally we find $S = 1$ or $S = 0$.

Example: Addition of three angular momentum for $s_1 = 1/2, s_2 = 1/2, s_3 = 1/2$. The procedure here is to first add the first two in the manner followed in the previous example, forming $S_{\text{int}} = 1$ or $S_{\text{int}} = 0$.

If $S_{\text{int}} = 1$, then this magnitude is larger than that of the third spin, and we thus choose an axis which is aligned along it. We therefore have $M_{\text{int}} = S_{\text{int}}$, and the possible projections of spin 3 on this axis are $m_{s,3} = \pm 1/2$. We then find that $S = S_{\text{int}} \pm 1/2$ and thus $S = 3/2$ or $S = 1/2$.

If $S_{\text{int}} = 0$, then the only possible value for the length of the total spin vector is $S = 1/2$.

For $S = 3/2$, we have $M_S = 3/2, 1/2, -1/2, -3/2$. For $S = 1/2$, we have $M_S = \pm 1/2$.

9.6 Optical transitions in many-electron atoms

Due to the high complexity of many-electron atoms, we will again not try to give a complete picture of how the atom interacts with optical fields. What is very useful is to have knowledge of which transitions are allowed to happen, and which ones are not. These are known as selection rules, as for the case of hydrogen.

Let us first remember the rules for the hydrogen atom in a magnetic field. We found previously that for a transition to have a non-zero probability of occurring, it must satisfy the condition

$$\Delta l = \pm 1 \quad (9.51)$$

where $\Delta l = l' - l$ is the change in the angular momentum magnitude quantum number. We also found that for an atom in a magnetic field, there is a selection rule on the projection of the angular momentum of the electron onto the field direction. This rule depended on the polarization of the light which is involved in the transition. We found that

$$\begin{aligned}\Delta m &= +1, \text{ right hand circular polarization} \\ \Delta m &= 0, \text{ linearly polarized along the field direction} \\ \Delta m &= -1, \text{ left hand circular polarization}.\end{aligned}$$

In a many electron atom, an important point with regards to optical transitions is that

- only one electron makes a transition.

For this case, then in order to consider the selection rules between states in a many-electron atom we need to consider how the change in the state of one of the electron (which is governed by hydrogen-like rules) will change the configuration, and also how it impacts on changes in the values of L , S and J . We will consider this in three parts, considering in turn the changes in quantum number at each level of approximation.

9.6.1 The configuration

If electron i makes a transition, then the selection rule is

$$\Delta l_i = \pm 1 \quad (9.52)$$

and therefore the quantum numbers of one of the electrons in the configuration of the atom must change. There is no selection rule for n_i so any change in n_i is possible. You should be careful to note that in the configuration *only one of the electrons* should change its state. It is exceedingly rare for two electrons to change state simultaneously.

9.6.2 The term

The term is governed by the values of S and L . Since the electric dipole operator is spatial, the spin is unaffected. The rule for S is thus that

$$\Delta S = 0. \quad (9.53)$$

For ΔL , we have to consider the ways in which a change of one unit of orbital angular momentum for a single electron impacts on the total orbital angular momentum. It is fairly

trivial to see that if the angular momentum vector for electron i is aligned with L , then if l_i changes length by one unit but stays in the same direction then L must also change by 1. This is true both for increases and decreases in l_i . We thus see that the value of L can change according to

$$\Delta L = \pm 1 . \quad (9.54)$$

There is another way in which the single electron can change state. This is a change for which $\Delta l_i = 1$ but $\Delta m_i = 0$. This can happen if the initial and final values of m_i are both zero and the total orbital angular momentum which has a magnitude given by the quantum number L lies along the z axis. In this case we can have

$$\Delta L = 0 \quad (9.55)$$

and thus we come to the selection rule for L which is

$$\Delta L = 0, \pm 1 \quad (9.56)$$

with the exception that

$$L = 0 \leftrightarrow L' = 0 . \quad (9.57)$$

This last condition comes about because if the sum of all the angular momentum vectors has zero length, changing the length of any of the vectors by any amount will make the sum non-zero.

9.6.3 The level

Since the spin does not change, but the magnitude and direction of the orbital angular momentum can, the selection rules for orbital angular momentum translate directly into those for the total angular momentum. We therefore have

$$\Delta J = 0, \pm 1 \quad (9.58)$$

with the exception that $\Delta J = 0$ is not possible if $J = 0$.