Lecture Notes on Quantum Physics

Matthew Foulkes
Department of Physics
Imperial College London

Spring and Summer Terms, 2011

Chapter 1

Introduction

1.1 What is Quantum Physics?

Quantum physics is a catch-all term for the ideas, devices and technologies made possible by the development of quantum mechanics in the early part of the 20th century. This course concentrates on the ideas behind quantum mechanics itself, but the broader field of quantum physics encompasses everything from the science of electronic devices and lasers to the philosophical mysteries of quantum measurement theory.

Quantum mechanics is our best current theory of matter and how it interacts. Matter in this context includes everything we normally think of as particles, waves, forces, and fields. In the quantum world, these are all (more or less) the same thing.

Given an experimental set-up, quantum mechanics tells you:

1. What can be measured.
2. The possible results of any measurement.
3. The probability of obtaining each of the possible results.

The rules used to calculate the probabilities, although abstract and mathematical, are precise and unambiguous. As a practical tool, quantum mechanics presents no difficulties and has been immensely successful.

1.2 Successes and Failures of Quantum Physics

To illustrate the extraordinary power and breadth of quantum theory, here are just a few of the phenomena it can explain:

Atomic structure and spectra
Properties and interactions of elementary particles
Semiconductor physics & devices
Superconductivity and superfluidity
The periodic table
Conductivity of copper
Hardness of diamond
Properties of neutron stars and white dwarfs
Magnetism
Radioactivity
Nucleosynthesis
Laser physics
Chemical reactions
Density of matter
Strength of steel
Stability of matter
Fisson/fusion
(The human brain?)

Some of the items in the list may strike you as classical, but if you ask one or two "why" questions you soon find yourself running in to quantum mechanics. Take the density of matter as an example: this depends on the size of an atom, which depends on the radius of an electron orbit and hence on quantum theory. In fact, the radius of a Hydrogen atom, known as the *Bohr radius* a_0, is given by

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} \approx 0.529 \times 10^{-10}\,\mathrm{m},$$

where $\hbar := h/2\pi \approx 1.05 \times 10^{-34}\,\mathrm{Js}$ is Planck's constant divided by 2π. The appearance of Planck's constant leaves no doubt that quantum theory is involved.

The version of quantum theory covered in this course neglects relativistic effects and is therefore an approximation, just as Newton's laws are an approximation to special relativity. The relativistic version of quantum mechanics, called quantum field theory, is very similar in outline but mathematically more difficult.

Quantum theory as a whole (including quantum field theory) has never been known to fail. Its applications have been limited by the difficulty of solving the equations, which are only tractable for rather simple systems, so there is no guarantee that problems will never be found; but even then quantum theory would remain useful, just as Newton's laws remained useful after the advent of special relativity. There is, as yet, no good quantised theory of gravity, but whether this indicates a fundamental problem with quantum mechanics or a failure of human ingenuity is unclear.

1.3 Quantum Weirdness

The most fascinating aspect of quantum mechanics is that it provides such a strange picture of the world. If you accept this picture — and given the practical successes of the theory it is difficult not to — you are left with no choice but to make fundamental changes to your idea of reality.

The first surprise is the *wave-particle duality* of the building blocks of matter. The world is not made of waves and particles, as in classical physics, but of peculiar hybrid objects with aspects of both. Suppose, for example, that you find an electron at $\mathbf{r}_1$ at time t_1 and then at $\mathbf{r}_2$ at a later time t_2. Since the electron is supposed to be a particle, you might imagine that it travelled along some specific path $\mathbf{r}(t)$ from $\mathbf{r}_1 = \mathbf{r}(t_1)$ to $\mathbf{r}_2 = \mathbf{r}(t_2)$. According to Feynman's path-integral formulation of quantum mechanics, however, this is wrong. In a precise mathematical sense (only hinted at in this course), the electron took all possible paths from $\mathbf{r}_1$ to $\mathbf{r}_2$ at once. Even worse, the components arriving along different paths interfered like waves.

Wave-particle duality is not the only strange aspect of quantum theory. The physical state of a quantum mechanical particle-wave is described by a *wave function*, $\psi(x,t)$, analogous to the amplitude of a classical wave. Unlike a classical wave, however, $\psi(x,t)$ does not evolve according to the classical wave equation,

$$\frac{\partial^2\psi(x,t)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi(x,t)}{\partial t^2} ,$$

where v is the phase velocity, but according to the time-dependent Schrödinger equation,

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial x^2} + V(x)\psi(x,t) = i\hbar\frac{\partial\psi(x,t)}{\partial t} ,$$

where m is the mass of the particle and $V(x)$ is the potential through which it moves. The most striking feature of Schrödinger's equation is that it has an i on the right-hand side, implying that the wave function is complex. Even if, by some fluke, $\psi(x,t)$ happened to be real at $t = 0$, it would not remain real. Complex waves are common in classical physics, of course, but the complex numbers are used only to simplify the mathematics and the physical waves remain real. In quantum theory, the wave function is *really* complex.

Perhaps the most puzzling aspect of quantum mechanics is that it predicts probabilities only. In classical physics, probabilities are used to describe our lack of knowledge of a physical system: if we know nothing about how a pack of cards has been shuffled, the probability of picking any particular card, say the three of spades, is $1/52$; if we know where all the cards are in advance, we can find the three of spades every time and there is no need for probability theory. Even for a complicated system such as the air in the Albert Hall, we could, in principle, measure the positions and velocities of all the molecules and predict the future evolution using Newton's laws; the probabilistic Maxwell-Boltzmann distribution is used only because the measurement is impractical and our knowledge incomplete.

It is tempting to imagine that the probabilistic nature of quantum theory arises in a similar way, and that quantum mechanics is just a rough statistical description of some more complicated underlying reality. As in the case of the air in the Albert Hall, we use a probabilistic description (there the Maxwell-Boltzmann equation; here quantum theory) only because our knowledge is incomplete. If we could discover the values of the *hidden variables* describing the underlying reality, we could dispense with probability theory altogether.

Hidden-variable theories are not completely impossible, but Bell's theorem shows that any such theory consistent with quantum mechanics must be non-local. This means, in effect, that every object in the universe has to be inter-dependent, and that we cannot interfere in one region without affecting everything else, no matter how far away. Most physicists find this idea so unsatisfactory that they prefer to think of nature as inherently probabilistic.

These ideas are fun, but the right time to think about them (if ever) is *after* you understand the workings of quantum theory. The aim of this course is to help you focus on the basics by making quantum mechanics as prosaic, straightforward and boring as possible! If you are unwilling to wait and want to find out more about the philosophical issues now, read *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy* by J.S. Bell. As well as inventing Bell's theorem and helping demystify the philosophical mess left by Bohr and friends, Bell (who was born in Belfast in 1928 and died in 1990) was a very good writer. His book is readable and quite accessible, requiring only a minimum of mathematics.

1.4 Course Content

This course covers the experimental evidence that led to the development of quantum mechanics and provides an introduction to quantum mechanical concepts and wave mechanics.

Concepts discussed include wave-particle duality, the wave function, the uncertainty principle, the Schrödinger equation, and the thorny question of measurement in quantum theory.

Schrödinger's wave mechanics is one of several equivalent formulations of non-relativistic quantum theory. The others, Heisenberg's matrix mechanics and Feynman's path-integral theory, look very different mathematically but describe the same physics and yield identical results.

To keep the mathematics as simple as possible, the introduction to wave mechanics in the second half of the course considers only a single non-relativistic particle in one dimension. The emphasis is on quantitative understanding and the practical application of physical principles rather than mathematical formalism (which is covered in some detail next year).

1.5 Some Useful Numbers

When does quantum mechanics matter? The conventional answer is at or below atomic/molecular length/energy scales. This section discusses some of the most important length and energy scales associated with everyday matter.
[A less conventional answer, to which I subscribe, and which is supported by the long list of quantum phenomena in Sec. 1.2, is that almost everything is quantum mechanical. In defence of this position, one of the classworks near the end of the course is about the quantum mechanics of a house brick.]

1.5.1 Lengths

Distance between air molecules

The ideal gas law

$$PV = Nk_BT \qquad \Rightarrow \qquad \frac{V}{N} = \frac{k_BT}{P} \; .$$

If, for simplicity, every molecule is assumed to occupy a cube of side a, so that $a^3 = V/N$, this gives

$$a^3 = \frac{k_BT}{P} \approx \frac{1.38 \times 10^{-23} \times 300}{1.01 \times 10^5}$$

and hence

$$a \approx 3.5 \times 10^{-9}\,\mathrm{m}.$$

Boltzmann's constant

The version of the ideal gas law taught in schools is

$$PV = n_m RT \ ,$$

where n_m is the number of moles and $R = 8.314\,\mathrm{JK}^{-1}$ is the gas constant. Since $n_m = N/N_A$, where N is the total number of molecules and N_A is Avogadro's number, this can be rewritten

$$PV = \left(\frac{N}{N_A}\right) RT = N\left(\frac{R}{N_A}\right) T = N k_B T \ .$$

The new constant $k_B := (R/N_A) \approx 1.38 \times 10^{-23}\,\mathrm{JK}^{-1}$ is known as Boltzmann's constant. On the whole, chemists prefer to work with moles and R, while physicists prefer molecules and k_B.

Distance between atoms in molecules/solids/liquids

A typical inter-atomic distance is a few $\times 10^{-10}$ m, otherwise known as a few Å. (1 Ångstrom := 10^{-10} m.)

Radius of an atom

A typical atomic radius is 1 Å.
(The radius of a Hydrogen atom, the Bohr radius a_0, is 0.529 Å.)

Radius of a nucleus

A typical nuclear radius is a few $\times 10^{-15}$ m.

1.5.2 Energies

Thermal energy at room temperature

$k_B T \approx 1.38 \times 10^{-23} \times 300 \approx 4.14 \times 10^{-21}$ J $\approx (1/40)$ eV.

The electron-volt

One eV is the kinetic energy gained by an electron falling through a potential difference of 1 V: $1\,\mathrm{eV} = qV = e \times 1\,\mathrm{Joules} = 1.6 \times 10^{-19}$ J.

Chemical bond

The energy of a typical covalent, ionic or metallic chemical bond is a few eV. (The van der Waals bonds between closed-shell atoms are much weaker.)

Binding energies of electrons in atoms

The energy required to strip an electron from an atom ranges from a few eV for the outermost “valence” electrons to thousands of eV for the innermost “core” electrons of heavy atoms.

Chapter 2

Light is Waves

2.1 Evidence for the Wave-Like Nature of Light

The waves with which we are most familiar — water waves, sound waves, the standing waves on a violin string — have several features in common.

Superposition and interference: If several waves overlap, the total displacement is the sum of the displacements of each.

Diffraction: Waves spread out after emerging from a narrow ($\lesssim \lambda$) opening.

Refraction: Waves change direction at boundaries between regions where the wave speed differs.

Light does all of these things, so light is a wave.

Historical note

The history of our understanding of light is interesting. Descartes and Newton, working in the 17th century, thought that light was a stream of particles, like bullets. It was not until the early 19th century that Thomas Young (born a Quaker in Somerset in 1773; learnt to read at 2; spoke a dozen languages; famous Egyptologist who helped decipher hieroglyphics; successful London physician) and others showed, apparently conclusively, that light was a kind of wave. For the next century or so, it was assumed that Newton and Descartes had been wrong. Following the arrival of quantum theory, it is now clear that Newton, Descartes and Young were all correct: light is both a particle and a wave.

Light waves are special in that they can travel through a vacuum and do not require a medium such as water or air. Most waves travel at a fixed speed relative to the medium that supports them, but light waves in vacuum have no medium and hence no preferred frame of reference. This in part explains why they always travel at a constant speed c relative to the observer. The constancy of the speed of light underlies special relativity but does not play an important role in this course. As far as we are concerned, light waves are much like any other waves.

The Schrödinger equation is a kind of wave equation and quantum mechanics is a theory of waves. To set the notation and establish a common starting point, the rest of this chapter revises some of the material from your Vibrations and Waves course.

2.2 Mathematical Description of Travelling Waves

2.2.1 Formula

The formula for a travelling wave is

$$\psi(x,t) = a\cos(kx - \omega t + \phi) \ . \tag{2.1}$$

At time $t = 0$, this wave is as shown in Fig. 2.1. Note that

$$a\cos(kx - \omega t + \phi) = a\cos(k[x - (-\phi/k)] - \omega t) \ ,$$

so there is a crest at $x = -\phi/k$ when $t = 0$.

As x increases by $2\pi/k$ at constant t, kx increases by 2π and $\psi(x,t)$ sweeps through one whole period. Hence

$$\left\{ \begin{array}{l} \lambda = 2\pi/k \ , \\ k = 2\pi/\lambda \ . \end{array} \right. \tag{2.2}$$

Similarly, as t increases by $2\pi/\omega$ at constant x, ωt increases by 2π and $\psi(x,t)$ sweeps through one period. Hence

$$\left\{ \begin{array}{l} T = 2\pi/\omega \ , \\ \nu = 1/T = \omega/2\pi \ , \\ \omega = 2\pi\nu \ . \end{array} \right. \tag{2.3}$$

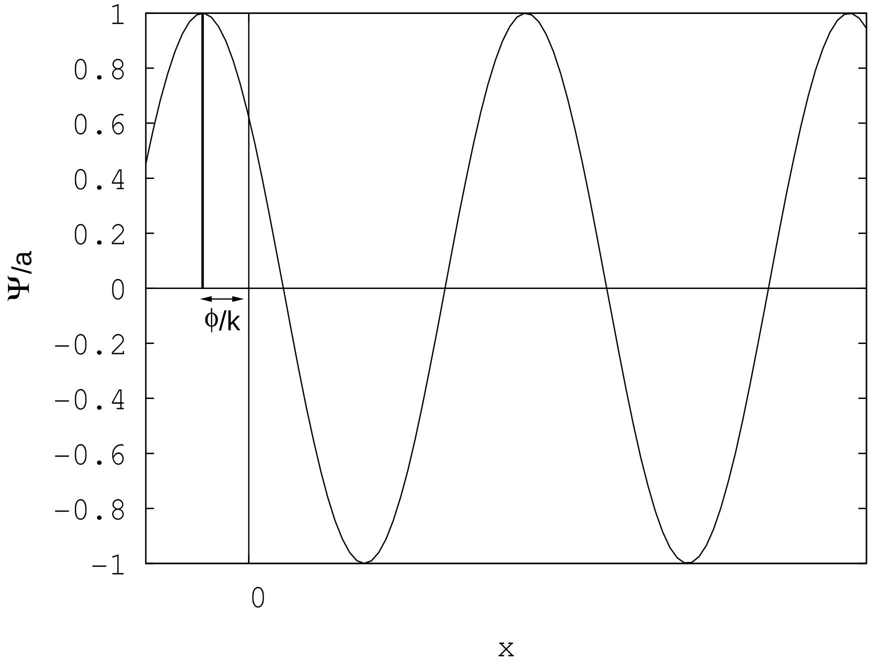


Figure 2.1: The travelling wave of Eq. 2.1 at time t=0.

2.2.2 Phase velocity

At time t, the wave has a crest where

$$kx - \omega t + \phi = 0$$

and hence where

$$x = -\phi/k + (\omega/k)\,t\ .$$

This shows that the *phase velocity* v_p, which is the velocity of the wave crests, is given by

$$v_p = \frac{\omega}{k} = \nu\lambda\ . \tag{2.4}$$

2.2.3 Group velocity

In quantum mechanics, we are often interested in *wave packets*, since these are the closest thing we can find to classical particles. A typical wave packet is shown

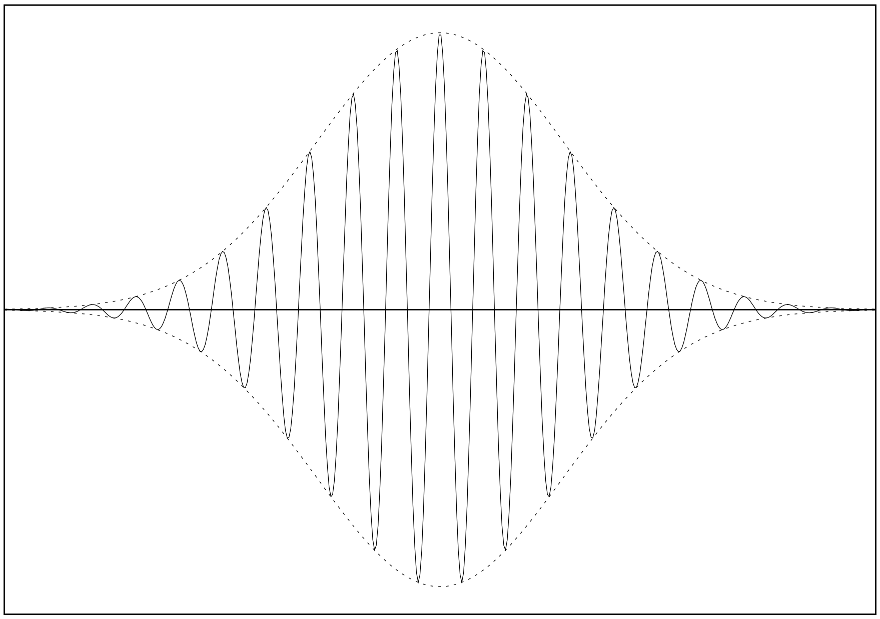


Figure 2.2: A wave packet.

in Fig. 2.2. Although the crests inside the wave packet move at the phase velocity v_p, the envelope of the packet — indicated by the dashed line in Fig. 2.2 — moves at the *group velocity*

$$v_g = \frac{d\omega}{dk} \; . \tag{2.5}$$

This equation was discussed in your Vibrations and Waves course but may not have been derived there. In case you are interested, a derivation is included in Chapter 6 of these notes.

As we shall see in Sec. 2.2.6, the phase velocity may be larger or smaller than the group velocity. If the phase velocity is larger than the group velocity, the crests within a wave packet travel more quickly than the envelope, appearing at the back, growing as they move forward, and then dying away at the front; if the phase velocity is smaller than the group velocity, the crests travel more slowly than the envelope, appearing at the front and dying away at the back.

Sign conventions

In this course, a right-going travelling wave is written as

$$\psi(x,t) = a\cos(kx - \omega t + \phi)\ .$$

Why not

$$\psi(x,t) = a\cos(\omega t - kx + \phi')\ ,$$

as in the Vibrations and Waves course? The two forms are equivalent if ϕ' is set equal to $-\phi$:

$$a\cos(\omega t - kx + \phi') = a\cos(\omega t - kx - \phi) = a\cos(kx - \omega t + \phi)\ .$$

(The final step used the fact that $\cos\theta = \cos(-\theta)$). Why, then, do we use one form in V&W courses and another in QM courses? The reason is historical: the inventors of quantum theory *chose* to write the time dependence of the wave function as "$-\omega t$" and built their choice into the form of the Schrödinger equation itself. It would be too confusing to change this convention now.

In fact, although the two forms are equivalent, one can argue that the "$+\omega t$" version used in V&W is better than the "$-\omega t$" version used in QM, because $\phi' = -\phi$ is a more natural definition of the phase shift. After all, as can be seen from Fig. 2.1, it is ϕ'/k $(= -\phi/k)$, not ϕ/k, that gives the position of the maximum at time $t = 0$.

2.2.4 Amplitude

The *amplitude* of a wave at a point x is the maximum displacement (of whatever it is that is waving) at that point. As shown in Fig. 2.3, the maximum displacement of a simple travelling wave,

$$\psi(x,t) = a\cos(kx - \omega t + \phi)\ ,$$

is equal to a at all points x.

2.2.5 Complex representation

Since $e^{i\theta} = \cos\theta + i\sin\theta$, the travelling wave $\psi(x,t) = a\cos(kx - \omega t + \phi)$ can be written as

$$\psi(x,t) \quad = \quad \text{Re}\left(ae^{i(kx-\omega t+\phi)}\right)$$

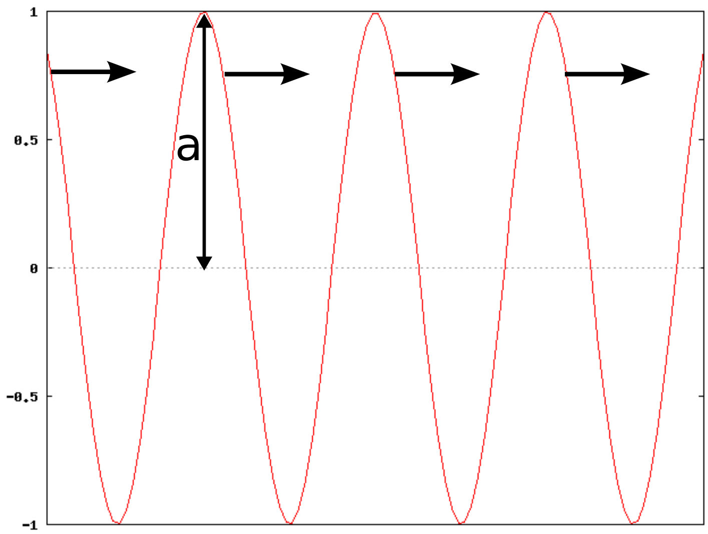


Figure 2.3: The crests of the travelling wave $a\cos(kx - \omega t + \phi)$ move steadily to the right, so the maximum displacement is a at all points x.

$$
\begin{aligned}
&= \text{Re}\left(ae^{i\phi}e^{i(kx-\omega t)}\right) \\
&= \text{Re}\left(Ae^{i(kx-\omega t)}\right)\ , \qquad (2.6)
\end{aligned}
$$

where

$$A = ae^{i\phi} \qquad (2.7)$$

is known as the *complex amplitude* of $\psi(x,t)$.

2.2.6 Dispersion relations

Any equation giving the angular frequency ω as a function of the wave vector k is called a *dispersion relation*.

Light: The dispersion relation for light is $\omega = ck$ (or, equivalently, $\nu = c/\lambda$). Hence

$$v_p = \frac{\omega}{k} = c\ , \qquad (2.8)$$

$$v_g = \frac{d\omega}{dk} = c\ . \qquad (2.9)$$

Since $v_p = v_g$, the crests within a wave packet move at the same speed as the envelope.

Quantum mechanical particle-waves: The dispersion relation for quantum mechanical particle-waves is $\omega = \hbar k^2/(2m)$. Hence

$$v_p = \frac{\omega}{k} = \frac{\hbar k}{2m} \, , \tag{2.10}$$

$$v_g = \frac{d\omega}{dk} = \frac{\hbar k}{m} \, . \tag{2.11}$$

Since $v_p < v_g$, the crests within a wave packet move more slowly than the envelope.

Large, gravity-dominated, deep-ocean waves: The dispersion relation for large ocean waves is $\omega = \sqrt{gk}$. Hence

$$v_p = \frac{\omega}{k} = \sqrt{\frac{g}{k}} \, ,$$

$$v_g = \frac{d\omega}{dk} = \frac{1}{2}\sqrt{\frac{g}{k}} \, .$$

Since $v_p > v_g$, the crests within a wave packet move more quickly than the envelope.

Small surface-tension-dominated water waves: The dispersion relation for small, surface-tension-dominated water waves is $\omega = \sqrt{\gamma k^3/\rho}$, where γ is the surface tension and ρ is the density. Hence

$$v_p = \frac{\omega}{k} = \sqrt{\frac{\gamma k}{\rho}} \, ,$$

$$v_g = \frac{d\omega}{dk} = \frac{3}{2}\sqrt{\frac{\gamma k}{\rho}} \, .$$

Since $v_p < v_g$, the crests within a wave packet move more slowly than the envelope.

[Bath-time experiments called for here!]

2.2.7 Intensity

Waves transmit energy. The energy density (energy per unit volume) at x is proportional to the *intensity*, defined here as the square of the amplitude, at that point. For example, if

$$\psi = a\cos(kx - \omega t + \phi) = \text{Re}\left(Ae^{i(kx-\omega t)}\right) \, ,$$

then

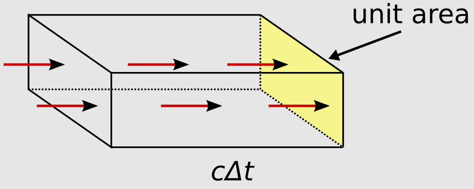
$$I = a^2 = ae^{i\phi}ae^{-i\phi} = AA^* = |A|^2 \, . \tag{2.12}$$

The intensity of a simple travelling wave is therefore independent of position and time. For more complicated waves and interference patterns, this is no longer the case.

Eyes and most optical instruments are sensitive to I and do not detect the phase ϕ directly.

Other definitions of intensity

In other courses, you may see intensity defined as the energy per unit volume, measured in Jm^{-3}, or the energy striking a unit area (oriented perpendicular to the beam) per second, measured in Wm^{-2}. To see how these two quantities are related, look at the following diagram showing a beam of light passing through a unit area.



In time Δt, all of the light energy in the box of length $c\Delta t$ passes through the right-hand face of unit area. Hence, the energy striking a unit area in time Δt is $uc\Delta t$, where u is the energy per unit volume in the box. Dividing by the time interval Δt gives the energy striking a unit area per second, which is equal to uc. Since $u \propto a^2$ and c is a constant, both u and uc are proportional to a^2. In this course, and in quantum physics in general, intensity always means simply a^2.

2.3 Interference

Because it is wave displacements that superpose, not intensities, the relative phases of the contributing waves matter:

pattern of relative phases (invisible) ⇒ pattern of intensities (visible)

In fact, if intensities added, there would be no interference. Section 2.2.7 showed that the intensity of a simple travelling wave is uniform, so no matter how many travelling waves were superposed, adding their intensities would give a uniform result.

The fact that intensities, and thus energy densities, do not add is somewhat strange. You might wonder, for example, whether the total energy is conserved when two travelling waves overlap and an interference pattern is formed. Fortunately, it turns out (see question 7 on problem sheet 1 for an example) that the position average of the intensity is always equal to the sum of the intensities of the contributing waves. The total energy is therefore correct, even though the formation of the interference pattern redistributes that energy over space.

The phenomenon of interference becomes even stranger in quantum theory, where $I(x,t)dx$ is the probability that a measurement of the position of a particle with wave function $\psi(x,t)$ yields a result between x and $x + dx$.

2.3.1 Example: the two-slit experiment

The two-slit interference experiment is the standard example used to help understand the meaning of the quantum mechanical wave function and will play an important role later in this course. In fact, according to Feynman, the two-slit experiment contains "the only mystery" of quantum theory. (If you would like to read about this now, the first few pages of the *Feynman Lectures on Physics: Quantum Mechanics v.3* are excellent.) To prepare for the later discussion, this section goes through the mathematics of the two-slit interference experiment for classical waves.

The set-up is as illustrated in Fig. 2.4. Suppose that the wave emerging from the upper slit travels a distance $\zeta_t = \zeta$ before hitting a distant screen. The wave emerging from the lower slit and hitting the same point on the screen travels a slightly longer distance, $\zeta_b = \zeta + d\sin\theta$. (This formula assumes that the screen is so far away that the rays from the two slits are effectively parallel; if the screen is close to the slits, the assumption of parallel rays is no good and the theory is harder.) Hence

$$\begin{aligned}
\psi(\zeta,t) &= Ae^{i(k\zeta_t-\omega t)} + Ae^{i(k\zeta_b-\omega t)} \\
&= Ae^{i(k\zeta-\omega t)} + Ae^{i(k(\zeta+d\sin\theta)-\omega t)} \\
&= A\left(1+e^{ikd\sin\theta}\right)e^{i(k\zeta-\omega t)}\ .
\end{aligned}$$

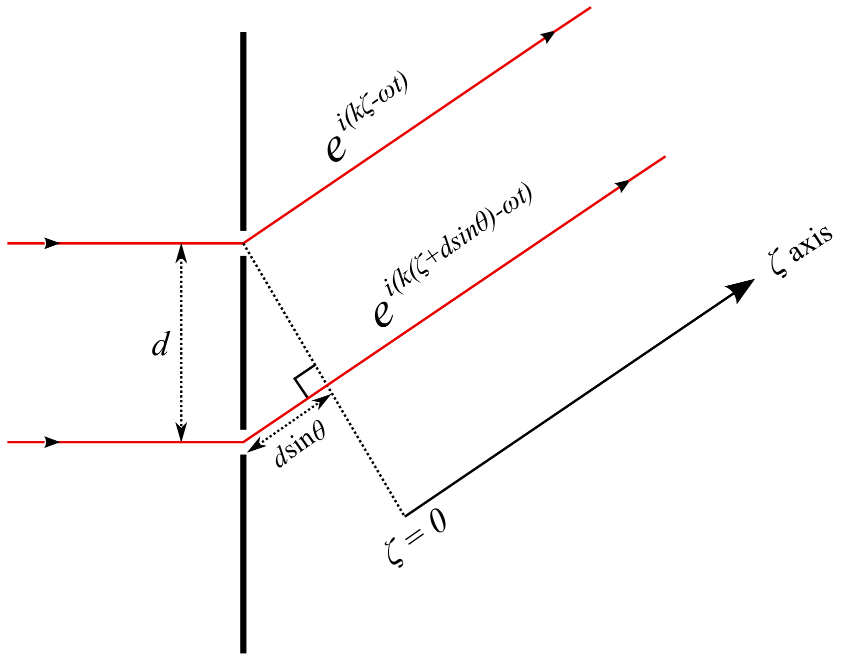


Figure 2.4: The two-slit experiment.

Introducing a new complex amplitude, $B = A\left(1 + e^{ikd\sin\theta}\right)$, this result becomes

$$\psi(\zeta, t) \quad = \quad Be^{i(k\zeta - \omega t)}\ .$$

The intensity emerging in the ζ direction is

$$\begin{aligned} I \quad = \quad |B|^2 \quad &= \quad A\left(1 + e^{ikd\sin\theta}\right) A^*\left(1 + e^{-ikd\sin\theta}\right) \\ &= \quad a^2\left(2 + e^{ikd\sin\theta} + e^{-ikd\sin\theta}\right) && (\text{because } A = ae^{i\phi}) \\ &= \quad 2a^2\left(1 + \cos(kd\sin\theta)\right) && \left(\text{because } \cos\alpha = (e^{i\alpha} + e^{-i\alpha})/2\right) \\ &= \quad 4a^2\cos^2\left(\frac{kd\sin\theta}{2}\right) && \left(\text{because } 1 + \cos\alpha = 2\cos^2(\alpha/2)\right). \end{aligned}$$

The diffraction pattern is as shown in Fig. 2.5. The first zero occurs where

$$\frac{kd\sin\theta}{2} = \frac{\pi}{2}$$

and hence where

$$d\sin\theta = \frac{\pi}{k} = \frac{\lambda}{2}\ .$$

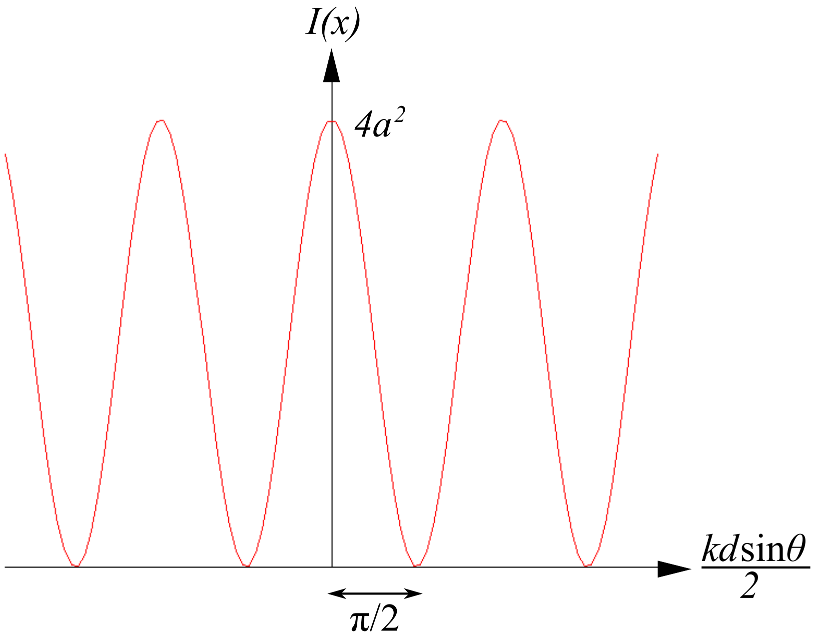


Figure 2.5: The two slit diffraction pattern.

Since $d\sin\theta$ is the path-length difference, this is exactly what one might have expected: the zero of the interference pattern occurs when the waves from the two slits are 180° out of phase.

The interference pattern obtained in a real two-slit experiment is more complicated because the slits are not infinitesimally wide. The $\cos^2[(kd\sin\theta)/2]$ oscillation is still visible, but its amplitude is modulated by an envelope, the shape of which corresponds to the diffraction pattern of a single slit of finite width. Question 8 of problem sheet 1 asks you to work out the diffraction pattern of a single slit.

Chapter 3

Light is Particles

3.1 Detecting Individual Photons

Light always arrives in “packets” called photons, the detection of which is nowadays routine. The *quantum efficiency* of a detector is the fraction of incident photons registered.

Figure 3.1 helps illustrate how quantum efficiency can be measured. The six electronically-generated panels illustrate the gradual formation of an image as increasing numbers of photons are collected by an idealised camera in which every photon is detected and produces a bright spot. Because photons arrive one by one, images made with low photon numbers look like collections of dots, with the overall pattern emerging gradually as the number of dots increases.

By shining a known number of photons into a real camera and comparing the image formed with pictures like these, one can establish the fraction of photons detected. If, for example, the image created with 9.3×10^5 photons looks like image C, which an ideal camera would produce with only 9.3×10^4 photons, the quantum efficiency is about 10%.

Similar experiments can be managed with the human eye, by asking someone to look at a projector emitting a known number of photons per second. If one assumes that the eye forgets about photons that arrived more than, say, $0.2\,\text{s}$ ago (which is about how long it takes for a human to register than an image has changed), one can convert the number of photons per second into a number of photons per image, and thus obtain a rough estimate of the quantum efficiency.

The results are quite surprising:

(i) The quantum efficiency of a human eye is about 5% at low light levels.

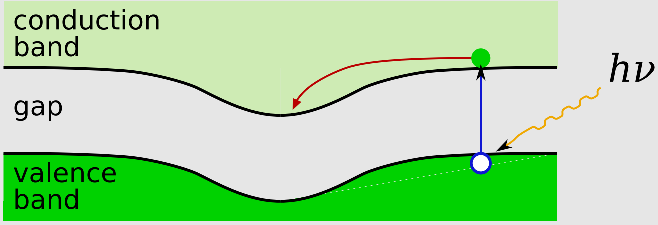
Figure 3.1: Images of a woman with flowers, as used by Rose in his article in *Physics Today* **42**, 24 (1989). The 6 images were made (electronically) using varying numbers of dots: A, 3×10^3; B, 1.2×10^4; C, 9.3×10^4; D, 7.6×10^5; E, 3.6×10^6; F, 2.8×10^7.

(ii) The quantum efficiency of photographic film is somewhat lower — typically less than 1%.

(iii) The quantum efficiency of a charge-coupled device (a "CCD", as found in a digital camera) ranges from 50–90%.

Charge-coupled devices

In a semiconductor, there is an energy gap between the filled electronic states, known as the valence band, and the empty electronic states, known as the conduction band. When a photon of sufficient energy is absorbed, it can excite an electron across this gap, out of the valence band and into the conduction band. In a CCD camera, voltages applied to conducting gates are used to apply a corrugated potential to the semiconductor, creating a huge array of tiny wells for electrons, several per pixel. As illustrated in the figure, an electron excited by a photon falls into the nearest well, where it sits, perhaps for hours, until it is read out. The read out is accomplished by applying voltages to the gates to make the stored electrons "click" over, well by well, to the edge of the camera, where they are collected and counted. For inventing the CCD at AT&T Bell Laboratories in 1969, Boyle and Smith were awarded (part of) the 2009 Nobel Prize in Physics.



A photon of energy $h\nu$ excites an electron from the valence band of a semiconductor to the conduction band. The electron then falls into the potential well to its left.

3.2 The Photoelectric Effect

The photoelectric effect is the emission of electrons when a beam of light strikes the surface of a solid. The photoelectric effect played an important role in the development of quantum mechanics and is still important today, although for different reasons. Photoemission is now used to investigate the nature of the electronic

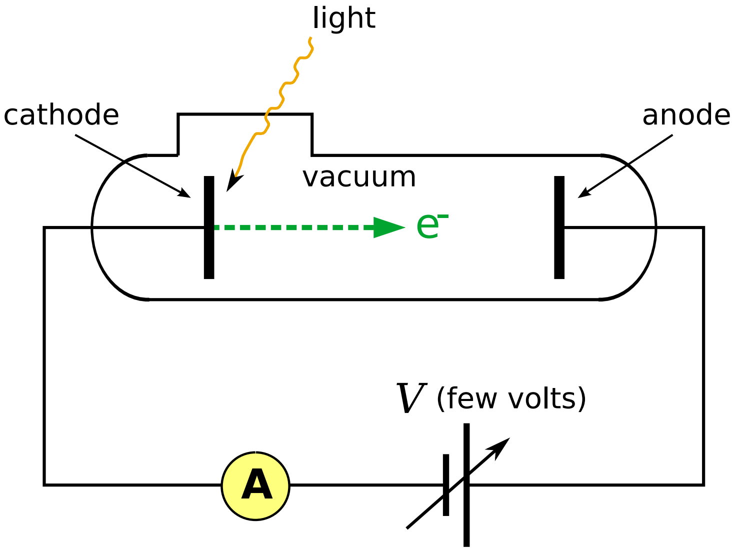


Figure 3.2: The apparatus used by Lenard in 1902 to observe the photoelectric effect.

energy levels inside solids. Inverse photoemission, in which photons are emitted when a beam of electrons strikes the surface, is also used.

A diagram of the apparatus Lenard used to observe the photoelectric effect in 1902 is shown in Fig. 3.2. A beam of light shines through a window into a vacuum chamber, where it hits a metallic cathode. The photons kick electrons out of the cathode (a *cathode* is any electrode that emits electrons), some or all of which move through the vacuum chamber to the *anode* (any electrode that absorbs electrons), causing a current to flow through the ammeter. By applying a voltage V between the anode and the cathode, the fraction of electrons collected can be altered. For simplicity, we assume that the cathode and anode are made of the same metal (photoemission happens for non-metals also, but the cathode and anode must conduct electricity if a current is to be measured). The dependence of the current on V at fixed photon wavelength is shown in Fig. 3.3.

Observations

(a) When V is large and positive, the current I saturates at I_{max}.

$$I_{\text{max}} \propto \text{light intensity}.$$

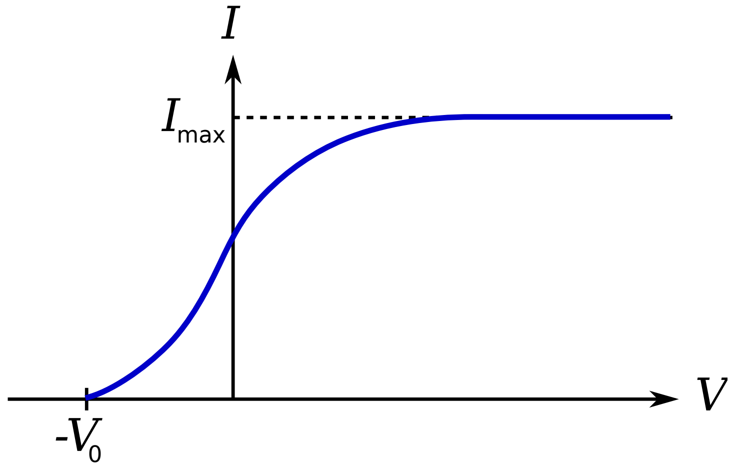


Figure 3.3: The current I measured by the ammeter as a function of the applied voltage V.

The obvious interpretation is that the current saturates when the electric field due to the applied potential V is strong enough to guide all of the photoemitted electrons to the anode, from where they flow back to the cathode through the ammeter. If every photoemitted electron is collected, increasing V cannot increase the current further.

(b) When $V < -V_0$, the current $I = 0$. The potential V_0 is known as the *stopping potential*.

The interpretation is that, at $V = -V_0$, only the most energetic photoemitted electrons are able to climb the potential barrier of height eV_0 to the anode.

$$\text{Maximum KE of emitted electrons} = \frac{1}{2}mv^2 = eV_0 \ . \tag{3.1}$$

Observations (a) and (b) have plausible classical explanations (although (b) is already strange: why such a sharp cut-off?). The following three observations are very hard to explain using classical physics.

(c) No delay is seen between the beginning of the illumination and the observation of the current.

The classical picture is that light energy arrives smoothly and the cathode warms up steadily until it becomes hot enough to emit electrons. One would not expect electron emission to begin immediately, especially when the light intensity is low.

(d) There is no light intensity threshold. The maximum current I_{max} is always proportional to the intensity, no matter how low the intensity.

In the classical picture, if the intensity were very low, one would not expect the cathode to warm up enough to emit electrons.

(e) The stopping potential V_0 depends on the metal and the photon frequency ν, but *not* on the light intensity.

According to classical physics, if the intensity were to increase, the cathode would get hotter and the emitted electrons would have more kinetic energy. The stopping potential would therefore increase.

The measured relationship relation between V_0 and ν is as shown in Fig. 3.4. The equation describing this graph is

$$eV_0 = h\nu - W \ . \tag{3.2}$$

The intercept $-W$ depends on the metal used for the cathode, but the slope $h \approx 6.63 \times 10^{-34}$ Js is universal, independent of cathode material and light intensity.

Einstein's simple but revolutionary explanation

- W, the *work function*, which is typically a few eV, is the minimum energy required to kick an electron out of the metal.
- Light arrives as packets/photons/quanta of energy $h\nu = \hbar\omega$.
- There is no delay or threshold intensity because one photon can kick out one electron.
- Figure 3.5 illustrates Einstein's idea and shows how the principle of conservation of energy leads to

$$h\nu - W = \frac{1}{2}mv_{\text{max}}^2 \ .$$

Since, from Eq. (3.1), $\frac{1}{2}mv_{\text{max}}^2 = eV_0$, it follows that

$$eV_0 = h\nu - W \ ,$$

which is Eq. (3.2).

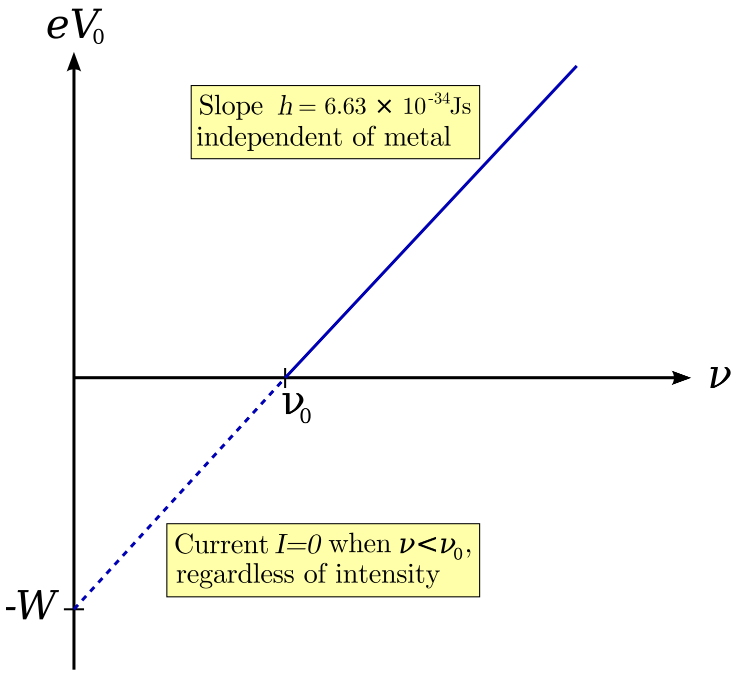


Figure 3.4: The relationship between the stopping potential V_0 and the photon frequency ν.

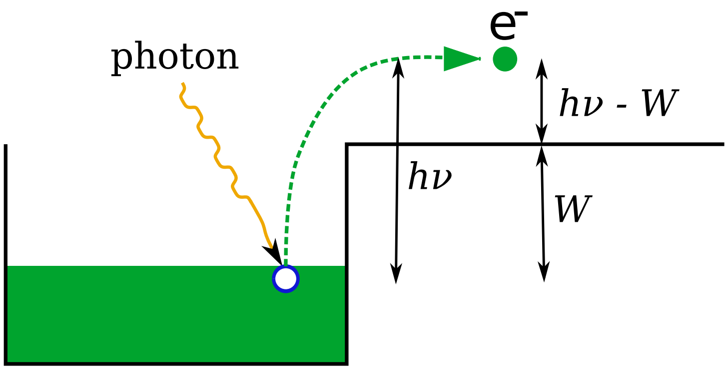


Figure 3.5: Einstein’s explanation of the photoeletric effect.

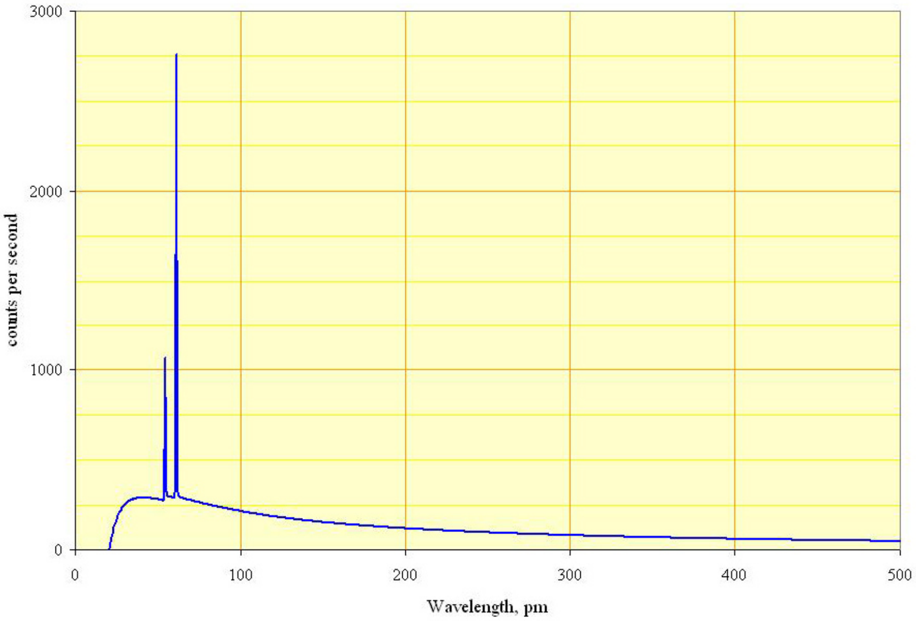


Figure 3.6: The spectrum of X-rays emitted when electrons of energy 60 keV strike a rhodium target (http://en.wikipedia.org/wiki/Bremsstrahlung).

3.3 The Production of X-Rays

X-rays, first observed by Röntgen in 1895, consist of penetrating electromagnetic radiation of wavelength 10^{-12} – 10^{-9} m. (For comparison, a typical visible photon has wavelength $\sim 500 \times 10^{-9}$m.) X-rays are emitted when electrons of energy 10^3 – 10^5 eV strike an anode made of atoms with a fairly high atomic number Z. The production of X-rays is the inverse of the photoemission process in which photons strike a solid and electrons are emitted, although the photon energies involved are much higher than those used in photoemission experiments.

The intensity spectrum of the X-rays produced when electrons of energy 60 keV strike a rhodium target (rhodium is a transition metal with $Z = 45$) is shown in Fig. 3.6.

- The spikes depend on the anode material and correspond to transitions between atomic energy levels (which will be discussed later in this course). The incoming electron knocks an electron out of an atomic core state, leaving a hole that is later filled by an electron dropping down from a higher atomic energy level. The energy released, which is exactly equal to the difference between the two atomic energy levels involved, is emitted as an X-ray photon.

- More interesting is the smooth background of bremsstrahlung (German for "braking radiation") emitted by the incoming electrons as they slow down after entering the anode. Maxwell's equations of electromagnetism predict that any accelerating/decelerating charge should emit radiation, so the existence of bremsstrahlung can be understood without invoking quantum theory. What cannot be understood is the sharp *material-independent* cut-off wavelength $\lambda_{\rm min}$, below which no radiation is emitted. Measurements show that $\lambda_{\rm min}$ is related to the energy E of the incoming electrons by

  $$\lambda_{\rm min} = \frac{hc}{E} \, ,$$

  where the constant $h \approx 6.63 \times 10^{-34}\,$Js is exactly as in the photoelectric effect. For the experiment shown in Fig. 3.6, where $E = 60\,$keV, this gives

  $$\lambda_{\rm min} = \frac{6.63 \times 10^{-34} \times 3.00 \times 10^{8}}{60 \times 10^{3} \times 1.60 \times 10^{-19}} \approx 2.07 \times 10^{-11}\,\text{m}.$$

  Since $c = \nu\lambda$, the minimum wavelength implies a maximum frequency, $\nu_{\rm max}$, given by

  $$\frac{c}{\nu_{\rm max}} = \frac{hc}{E}$$

  and hence

  $$h\nu_{\rm max} = E \, .$$

  The quantum mechanical explanation of the wavelength cut-off is obvious: an electron of energy E cannot produce a photon of energy $h\nu > E$. (The work function W is too small to be noticeable in these very high-energy experiments.)

3.4 Planck's Constant and Black-Body Radiation

Planck introduced his new constant h in 1900, before any of the experiments mentioned so far, as part of an attempt to understand black-body radiation (BBR). The theory of BBR is covered in detail in the second-year statistical physics course, but the subject is so interesting that I wanted to include a preview here. This section of the notes is not examinable.

What is black-body radiation?

Imagine a well-insulated oven with a very small hole from which electromagnetic radiation can leak out. The photons inside the oven make many collisions with the walls before escaping through the hole and have plenty of time to reach thermal equilibrium at the oven temperature T. Just as with the Maxwell-Boltzmann distribution of molecular speeds in a gas, the radiation in the oven (and hence the radiation emitted) is found to have a *universal* spectrum that depends only on the temperature of the oven. Expressed mathematically, this means that the energy per unit volume in the frequency range ν to $\nu + d\nu$ can be written as $U(\nu, T)d\nu$, where the function $U(\nu, T)$ is independent of the material from which the oven is made or how it is constructed. To the extent that other hot objects act like black bodies (and many do), this explains why you can tell the temperature from the colour of the glow alone, without knowing what the object is made of.

Why is a glowing oven called a black body? Photons entering the oven via the tiny hole have to make so many collisions before re-emerging that they are almost certain to be absorbed first. The (hole in) the oven is therefore an ideal absorber as well as an ideal emitter.

Black-body radiation and the second law of thermodynamics

The fact that all black bodies of the same temperature emit identical radiation seems surprising at first. One way to see that it must be true is to use the second law of thermodynamics, which says, more or less, that there is no free lunch. More precisely, the Clausius formulation of the second law states that heat cannot flow spontaneously from a colder to a hotter object. If it did, the resulting temperature difference could be used to drive a heat engine and we could get unlimited energy for nothing.

How does the second law relate to BBR? Suppose that we were able to find two ovens, A and B, with different black-body spectra. We could then choose a frequency ν and temperature T at which the energy density in A was greater than the energy density in B. If we connected the two ovens together via a filter that only passed electromagnetic radiation of that frequency, as shown in Fig. 3.7, there would be a net energy flow from oven A to oven B. A temperature difference would be generated spontaneously, heat would flow from the colder oven to the hotter oven, and the world's energy problems would be solved.

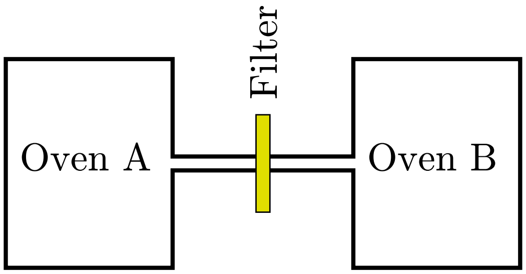


Figure 3.7: Two ovens joined by a narrow tube with a frequency selective filter.

Examples of black-body radiation

- Heat from an oven or fridge open a crack.
- Photons created deep within the sun are scattered huge numbers of times and have plenty of opportunity to reach thermal equilibrium before emerging. The sun is an almost perfect black-body emitter at about 6000K.
- The cosmic microwave background is radiation left over after the decoupling of matter and radiation in the early universe. Although the cosmic microwave background has been cooled by the expansion of the universe since decoupling, its spectrum still has an almost perfect black-body form.

Theoretical difficulties

Physicists at the end of the 19th century thought they knew how to calculate the spectrum of BBR. As illustrated in Fig. 3.8, electromagnetic standing waves in a cavity can have various wavelengths (and hence frequencies $\nu = c/\lambda$), which are set by the cavity size. The possible standing waves can be listed and the number $N(\nu)d\nu$ with frequencies in the range ν to $\nu + d\nu$ worked out. (You will go through this calculation next year.) For a large enough cavity, the result is that

$$N(\nu)d\nu = \frac{8\pi\nu^2 d\nu}{c^3}$$

per unit volume of the cavity.

According to the classical theory of electromagnetism, every standing wave acts like a simple harmonic oscillator, and the mean energy of a classical harmonic oscillator at temperature T is known to be kT. Thus, the energy spectrum ought

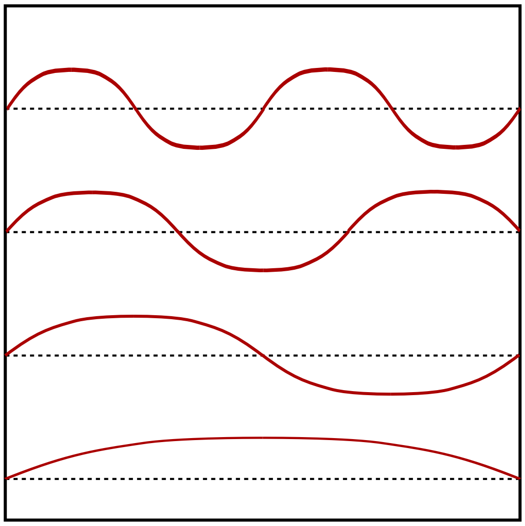


Figure 3.8: Electromagnetic standing waves in a cavity.

to be

$$U(\nu,T)d\nu \;=\; kT \times \frac{8\pi\nu^2}{c^3}d\nu\,.$$

This function, known as the Rayleigh-Jeans spectrum, is shown in Fig. 3.9, compared with the experimentally measured spectrum. The agreement is good at low frequencies, but the Rayleigh-Jeans formula predicts that the energy per unit frequency carries on rising as the frequency rises. Since there is no upper limit on the frequency (no lower limit on the wavelength of the waves that fit into the box), the total energy per unit volume is infinite! This theoretical disaster was christened the *ultra-violet catastrophe*.

Planck was able to fix the problem and reproduce the experimentally observed spectrum exactly by making the ad-hoc assumption that the walls of the cavity and the electromagnetic standing waves within the cavity could only exchange energy in chunks of size $h\nu$. (Nowadays, of course, we call these chunks photons.) If we accept this assumption, which much have seemed pretty wild at the time, the possible energies E of the standing wave of frequency ν are

$$0,\;\; h\nu,\;\; 2h\nu,\;\; 3h\nu,\;\; \ldots.$$

Actually, this is slightly wrong: we now know that the allowed energies are $(n+\frac{1}{2})h\nu$, where $n = 0,1,2,\ldots$ and the extra $\frac{1}{2}h\nu$ is the *zero-point energy*.

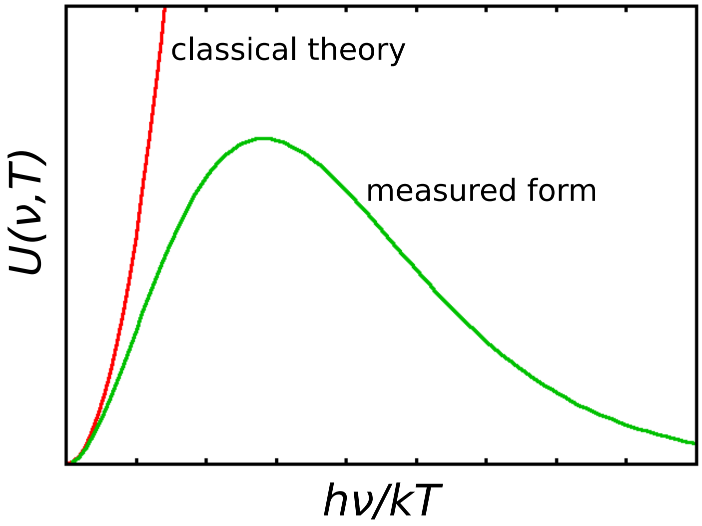


Figure 3.9: Comparison of the measured black-body spectrum and the classical Rayleigh-Jeans result, $U(\nu, T) = 8\pi\nu^2 kT/c^3$.

Fortunately, as you can check for yourself, the addition of the zero-point energy does not affect the form of the function $U(\nu, T)$ derived below.

Planck knew from classical statistical mechanics that the probability that a physical system at temperature T has energy E is $Ce^{-E/kT}$, where C is a constant. The value of C can be found from the condition that the system always has *some* energy, so the probabilities must add up to one:

$$1 = C \sum_{n=0}^{\infty} e^{-nh\nu/kT} \ .$$

The sum is a geometric series with first term $a = 1$ and ratio $r = e^{-h\nu/kT}$. Since $\sum_{n=0}^{\infty} ar^n = a/(1-r)$, it follows that $C = (1-r)/a = 1 - e^{-h\nu/kT}$. The probability p_n that the standing wave of frequency ν has energy $nh\nu$ is therefore

$$p_n = (1 - e^{-h\nu/kT})e^{-nh\nu/kT} \ .$$

The *mean* energy $\langle E_\nu \rangle$ in the standing wave of frequency ν is

$$\langle E_\nu \rangle = \sum_{n=0}^{\infty} (nh\nu)p_n = h\nu(1 - e^{-h\nu/kT}) \sum_{n=0}^{\infty} ne^{-nh\nu/kT} \ .$$

This series can also be evaluated (hint: $\sum_{n=0}^{\infty} ne^{-nx} = -\frac{d}{dx}\sum_{n=0}^{\infty} e^{-nx}$), giving

$$\langle E_\nu \rangle = h\nu(1 - e^{-h\nu/kT}) \times \frac{e^{-h\nu/kT}}{(1 - e^{-h\nu/kT})^2} = \frac{h\nu}{e^{h\nu/kT} - 1} \; .$$

Since the number of standing waves per unit volume is $8\pi\nu^2 d\nu/c^3$ and the energy per standing wave is $\langle E_\nu \rangle$, the total energy in the frequency range ν to $\nu + d\nu$ per unit volume of the cavity must be

$$U(\nu, T)d\nu \;\; = \;\; \frac{8\pi\nu^2 d\nu}{c^3} \times \frac{h\nu}{e^{h\nu/kT} - 1} \; .$$

If h is set to 6.63×10^{-34} Js, this form fits the experimentally measured spectrum exactly.

At very high temperatures, when $kT \gg h\nu$,

$$\frac{h\nu}{e^{h\nu/kT} - 1} \approx kT$$

and

$$U(\nu, T)d\nu \approx \frac{8\pi\nu^2 kT d\nu}{c^3},$$

as in the Rayleigh-Jeans law. At low temperatures, however, when $kT \ll h\nu$, the atoms in the walls do not have enough energy to make a photon of energy $h\nu$ and the electromagnetic standing wave of frequency ν is not excited. This explains why the Planck formula for the energy density drops to zero at high frequency, fixing the ultra-violet catastrophe.

(In the Structure of Matter course, you will learn that the vibrations of oxygen and nitrogen molecules do not contribute to the heat capacity of air at room temperature because the quantum of energy required to excite the vibrations is $\gg kT$. The similarity to Planck's explanation of BBR is obvious.)

Planck's achievement was remarkable, but it is not clear that he took the idea of quantisation as far as Einstein. He understood that the mechanism of energy exchange between the oven walls and the electromagnetic standing waves was chunky, but did not seem to realise that this might be because of the chunkiness of the electromagnetic field itself.

3.5 The Planck and De Broglie Equations

The last few sections of these notes were meant to convince you that photons are particles with energy

$$E = h\nu = \hbar\omega \qquad \text{(Planck's equation)} \tag{3.3}$$

Special relativity tells us that the energy and momentum of a relativistic particle are related by

$$E^2 = p^2c^2 + m^2c^4 \ ,$$

where m is the rest mass. Since photons have $m = 0$, it is reasonable to guess that their energy and momentum are related by

$$E^2 = p^2c^2 \qquad \Rightarrow \qquad E = pc \ . \tag{3.4}$$

(Strictly, of course, this should be $E = \pm pc$, since E is always postive but p is negative if the photon is moving in the $-x$ direction.)

Students sometimes argue that, since momentum is mv and photons have $m = 0$, photons cannot have momentum. This argument is wrong because the Newtonian formula $p = mv$ only applies when $v \ll c$. For relativistic massive particles, the definition is

$$p = m\gamma v = \frac{mv}{\sqrt{1 - v^2/c^2}} \ .$$

Photons have $m = 0$ and $v = c$, so the numerator and denominator are both zero and the ratio is undefined. This leaves open the possibility that photons may have momentum even though they do not have rest mass. The formula $E = pc$ derived above from Planck's law can also be obtained directly from the classical theory of electromagnetism and was known long before the arrival of quantum theory.

Combining Eqs. (3.3) and (3.4) gives

$$h\nu \ = \ pc \qquad \Rightarrow \qquad p \ = \ \frac{h\nu}{c}$$

and hence

$$p = \frac{h}{\lambda} = \hbar k \qquad \text{(The de Broglie equation)} \tag{3.5}$$

In fact, as we shall see in Chapter 4, it turns out that Eqs. (3.3) and (3.5) apply to *all* particles, not just photons.

Some of the original experimental evidence in favour of the hypothesis that photons have momentum $p = h/\lambda$ is discussed in the next section of these notes. Compelling recent evidence is obtained every day at CERN, where the tracks of

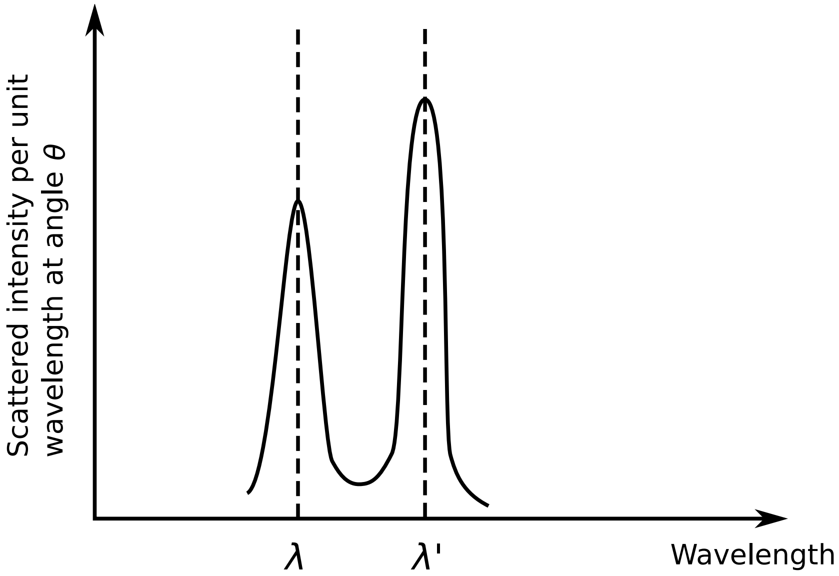


Figure 3.10: The intensity of Compton-scattered X-rays per unit wavelength as a function of wavelength.

photons created in particle collisions are measured and the momentum conservation law does not work unless the photon momentum is taken into account. Another strand of evidence is discussed in Classwork 2, which describes the "optical molasses" technique used to cool atoms in magnetic traps. More generally, whenever photons are absorbed or reflected by an object, momentum is transferred and the object feels a force. Radiation therefore exerts a pressure, which is small but measureable.

3.6 Compton Scattering

In 1923, when X-ray physics was cool, Compton decided to study the scattering of X-rays ($\lambda = 0.071\,\mathrm{nm}$, $\nu = c/\lambda = 4.23 \times 10^{18}\,\mathrm{Hz}$, $E = h\nu = 17.5\,\mathrm{keV}$) from graphite. According to classical physics, X-rays are electromagnetic travelling waves of frequency $\nu = c/\lambda$. The fields associated with these waves exert forces on the electrons, causing them to vibrate at the same frequency. Since oscillating charges emit radiation, Compton expected to see scattered radiation at frequency ν and perhaps also a few harmonics such as 2ν or 3ν.

What he actually saw is illustrated in Fig. 3.10. As well as the scattered ra-

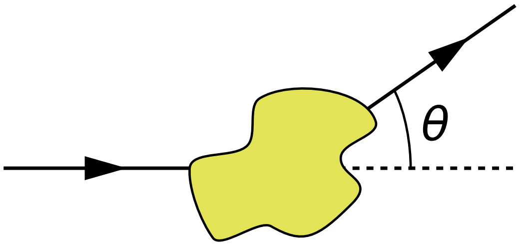


Figure 3.11: The Compton scattering angle.

diation of frequency ν and wavelength λ predicted by classical theory, he also saw scattered X-rays of longer wavelength λ' (lower frequency ν'). It was like shining a blue light at a mirror and seeing it come back red! The wavelength shift depended on the scattering angle θ illustrated in Fig. 3.11. In fact, careful measurements showed that

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta) , \tag{3.6}$$

where m is the mass of an electron.

Compton found a simple and convincing explanation of his results based on the following assumptions:

- X-rays scatter from electrons (hence the appearance of the electron mass m in Eq. (3.6)).

- Since the X-ray photon energy $h\nu$ is much greater than the binding energy of an electron in a carbon atom, it suffices to consider scattering from a free electron.

- X-rays arrive as photons of energy $E = pc = h\nu$ (which was by then quite well established) and momentum $p = h/\lambda$ (which was still controversial).

A typical Compton scattering event is shown in Fig. 3.12. A photon of momentum $\mathbf{p}$ and energy $h\nu = pc$ hits a stationary electron of mass m. After the collision, the photon moves off with momentum $\mathbf{p}'$ and energy $h\nu' = p'c$ at angle θ, and the electron recoils with momentum $\mathbf{P}$ and energy $E = \sqrt{m^2c^4 + P^2c^2}$. Since the energy transferred to the electron may be large, the relativistic relationship between the electron energy and momentum has been used.

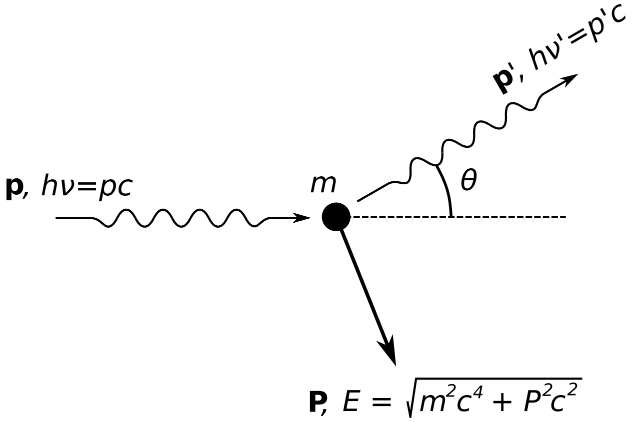


Figure 3.12: A photon of momentum $\mathbf{p}$ and energy $h\nu = pc$ hits a stationary electron of mass m. After the collision, the photon moves off with momentum $\mathbf{p}'$ and energy $h\nu' = p'c$ at angle θ, and the electron recoils with momentum $\mathbf{P}$ and energy $E = \sqrt{m^2c^4 + P^2c^2}$.

Applying the principle of energy conservation to the collision gives

$$\begin{aligned} pc + mc^2 &= p'c + \sqrt{m^2c^4 + P^2c^2} \\ \left[(p - p')c + mc^2\right]^2 &= m^2c^4 + P^2c^2 \;. \end{aligned} \tag{3.7}$$

Momentum conservation gives

$$\mathbf{p} = \mathbf{p}' + \mathbf{P}$$

and hence

$$\begin{aligned} P^2 &= (\mathbf{p} - \mathbf{p}') \cdot (\mathbf{p} - \mathbf{p}') \\ &= p^2 - 2\mathbf{p} \cdot \mathbf{p}' + p'^2 \\ &= p^2 - 2pp' \cos\theta + p'^2 \\ &= (p - p')^2 + 2pp'(1 - \cos\theta) \;. \end{aligned} \tag{3.8}$$

Using Eq. (3.8) to eliminate P from Eq. (3.7) gives

$$(p - p')^2c^2 + 2mc^3(p - p') + m^2c^4 = m^2c^4 + (p - p')^2c^2 + 2pp'c^2(1 - \cos\theta)$$

and hence

$$2mc^3(p - p') = 2pp'c^2(1 - \cos\theta)$$

$$\begin{aligned}
\Rightarrow \qquad \frac{1}{p'} - \frac{1}{p} &= \frac{1}{mc}(1 - \cos\theta) \qquad \text{(dividing through by } 2mc^3pp'\text{)} \\
\Rightarrow \qquad \lambda' - \lambda &= \frac{h}{mc}(1 - \cos\theta) \qquad \text{(since } p = h/\lambda\text{),}
\end{aligned}$$

exactly as observed.

Compton's theory shows that an X-ray scattered from a free electron increases its wavelength from λ to λ', explaining the longer-wavelength peak in Fig. 3.10. Sometimes, however, the incoming X-ray fails to knock the target electron out of its atom and momentum is transferred to the atom as a whole. Since atoms are much more massive than electrons, the energy transferred in such collisions is almost zero, explaining the peak of scattered X-rays at the incident wavelength λ.

Chapter 4

Particles are Waves

4.1 The Planck and De Broglie Equations Revisited

Louis de Broglie's 1924 hypothesis was much bolder than explained so far. He proposed, with essentially no supporting evidence, that *all* particles, not just photons, have wave-like properties and satisfy

$$p = h/\lambda = \hbar k \qquad \text{(The de Broglie equation)} \tag{4.1}$$

$$E = h\nu = \hbar\omega \qquad \text{(The Planck equation)} \tag{4.2}$$

Even more surprising is that he made this proposal in his PhD thesis! Few physicists, including his PhD examiner Paul Langevin, the inventor of Langevin dynamics and the Langevin equation (both of which I use in my own work), believed de Broglie's outlandish proposal at the time. Later, de Broglie said that Langevin was "probably a little stunned by the novelty of my ideas." In fact, Langevin was so worried that he wrote to Einstein to ask whether de Broglie should be allowed to pass. Einstein seems to have had doubts too, but fortunately for his future reputation said yes. After some reflection, Einstein also said: "I believe it (de Broglie's work) is a first feeble ray of light on this the worst of our physics enigmas." Davisson and Germer demonstrated electron diffraction experimentally in 1927 and de Broglie won a Nobel Prize in 1929, only five years after finishing his PhD.

4.2 Dispersion Relation of Particle Waves

For the time being, let us simply assume that the Planck and de Broglie equations, $E = h\nu = \hbar\omega$ and $p = h\lambda = \hbar k$, apply to particles such as electrons as well

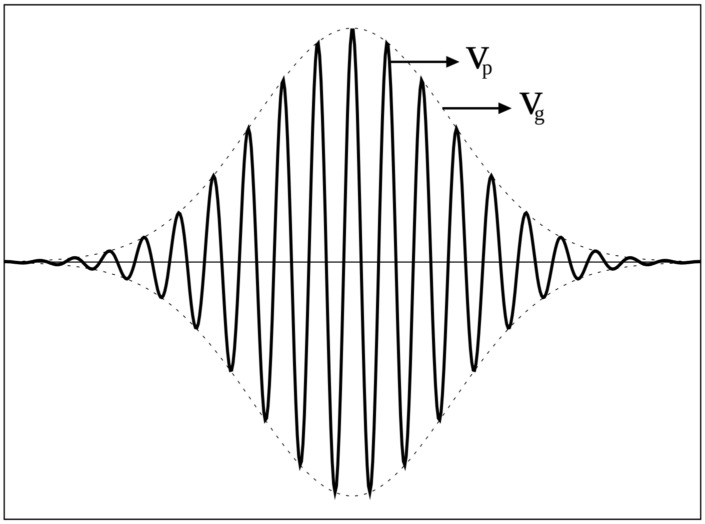


Figure 4.1: The phase and group velocities of a wave packet.

as photons. (Some of the experimental evidence in favour of this assumption is discussed in Sec. 4.4.) For non-relativistic massive particles, we also know that

$$E = \frac{1}{2}mv^2 = \frac{p^2}{2m} \; .$$

Hence, the dispersion relation of the de Broglie particle-waves must be

$$\hbar\omega = \frac{(\hbar k)^2}{2m}$$

$$\omega = \frac{\hbar k^2}{2m} \tag{4.3}$$

The phase velocity (the speed of the wave crests — see Fig. 4.1) is

$$v_p = \frac{\omega}{k} = \frac{\hbar k}{2m} \; ,$$

and the group velocity (the speed of the envelope of a wave packet — see Fig. 4.1) is

$$v_g = \frac{d\omega}{dk} = \frac{\hbar k}{m} = \frac{p}{m} \; .$$

This velocity p/m of the "particle" is the same as group velocity of the wave packet, which makes sense.

4.3 Examples

- **An electron in an atom, molecule or solid** has

$$E = \frac{p^2}{2m} = \frac{h^2}{2m\lambda^2}$$

and hence

$$\lambda = \frac{h}{\sqrt{2mE}} \, . \tag{4.4}$$

If $E = 5\,\mathrm{eV} = 5 \times 1.60 \times 10^{-19}\,\mathrm{J}$, which is typical for valence electrons in ordinary matter, then

$$\lambda \approx 5.5 \times 10^{-10}\,\mathrm{m}.$$

This is larger than the radius of an atom and comparable to or larger than the spacing between atoms in molecules and solids, so the behaviour of the electrons is very wave-like. The electron waves diffract from the atoms, refract from changes in potential, and interfere just like light waves. The quantum theory of atomic bonding (and chemistry in general) is essentially a wave theory.

Electron energy scales

Why is the typical energy $E \sim 5\,\mathrm{eV}$ of a valence electron in an atom, molecule or solid so much larger than kT (which is only about $1/40\,\mathrm{eV}$ at room temperature)? One way to explain this is by analogy: the speeds at which planets orbit the sun have nothing to do with their temperature, but are determined by Newton's laws, $mv^2/r = GMm/r^2$. In a similar way, the speeds at which electrons "orbit" nuclei are determined by the quantum mechanical laws of motion. Thermal excitations increase the average speed a little, in metals at least, but the change is relatively tiny. Another contributing factor is the Pauli exclusion principle, which says that only one electron can occupy any quantum state. This makes it impossible to place all of the electrons in low-energy atomic orbitals.

- **An N_2 molecule in air at 300K** has

$$\begin{aligned} m_{\mathrm{N}_2} &= 2 \times 14 \times 1.66 \times 10^{-27}\,\mathrm{kg}\ , \\ E &= \frac{3}{2}kT \quad \approx \quad 0.039\,\mathrm{eV}\ , \\ \lambda &= \frac{h}{\sqrt{2m_{\mathrm{N}_2}E}} \quad \approx \quad 2.8 \times 10^{-11}\,\mathrm{m}. \end{aligned}$$

Since the wavelength is much smaller than the distance between molecules (which is $\sim 3.5 \times 10^{-9}\,\mathrm{m}$ — see Sec. 1.5), it is easy to make N_2 wave packets much smaller than the inter-molecular separation. These tiny wave packets move around and collide almost like particles, explaining why it makes sense to view an air molecule (treated as a whole, ignoring the behaviour of the electrons inside it) as a classical object in most cases. Quantum (wave-like) effects become more important when the temperature is very low and the de Broglie wavelength larger.

More surprising is that the de Broglie wavelength of an N_2 molecule is smaller than a single atom. This means that atoms and molecules (again regarded as whole objects) in solids can also be treated as classical particles in most cases.

How can the de Broglie wavelength of a molecule be smaller than the molecule itself? The answer is that the wavelength obtained by treating the molecule as a single entity (rather than a collection of electrons and nucleons, each of which has its own de Broglie wave) is associated with the molecule's centre of mass. The molecule itself can easily be larger than the uncertainty in its centre-of-mass position.

- **A person** has

$$\begin{aligned} m &\sim 60\,\mathrm{kg} \\ v &\sim 1\,\mathrm{m\,s^{-1}} \qquad \text{(say)} \\ \lambda &= \frac{h}{p} \sim 1.1 \times 10^{-35}\,\mathrm{m!} \end{aligned}$$

The QM/wave-like properties of macroscopic objects are almost always undetectable.

4.4 Evidence that Particles are Waves

Electron diffraction

In 1927, Davisson and Germer showed that electrons can be diffracted by the regular array of atoms in a crystal, which acts much like a diffraction grating. This experiment is described in classwork 3.

Because electrons are charged and not very heavy, they are strongly scattered by solids and penetrate only very small distances. This makes electron diffraction a useful (and widely used) technique for studying how atoms are arranged at surfaces. It turns out that the 2D crystal structures of surfaces often have rather little to do with the 3D crystal structures inside solids.

Thermal neutron diffraction

A neutron ($m \approx 1.67 \times 10^{-27}\,\mathrm{kg}$) with energy $E = 3kT/2$ at $T = 300\,\mathrm{K}$ has $\lambda = h/\sqrt{2mE} \approx 1.5 \times 10^{-10}\,\mathrm{m}$, which is comparable to the size of an atom. Just as with electrons, therefore, beams of neutrons are diffracted by the regular arrays of atoms in crystals.

Neutrons are much heavier than electrons and uncharged, so they penetrate deep inside solids and diffract from the bulk crystal structure, not the surface. This makes neutron diffraction a useful alternative to X-ray diffraction. Because neutrons have a magnetic moment, the diffraction pattern depends on the arrangement of magnetic moments in the target, making neutron diffraction especially useful for investigating magnetic solids.

He atom diffraction

The mass of a ^{4}He atom is about four times that of a neutron, so beams of low-energy He atoms also diffract from crystals. Unlike neutrons, however, low-energy He atoms are large enough to find it very difficult to penetrate solids. He diffraction, like electron diffraction, is a surface sensitive technique.

Scanning tunnelling microscopy

The scanning tunnelling microscope (STM) was invented by Gerd Binnig and Heinrich Rohrer at IBM Zürich in 1981. Although the invention was to some extent serendipitous (they were trying to build a microscope but had no idea they would achieve atomic resolution), their work was so obviously important that they were awarded the Nobel prize just a few years later, in 1986.

A schematic diagram of an STM is shown in Fig. 4.2. A tiny metallic tip (an atom or two across at its end) is scanned horizontally a few Å above a conducting surface. A small voltage is applied between the tip and the surface, causing electrons to jump across the gap. The jumping process, which is known as quantum mechanical tunnelling (and will be discussed right at the end of this course), is exquisitely sensitive to distance, so the current rises and falls as the tip is scanned across the surface and approaches or moves away from the nearest surface atom.

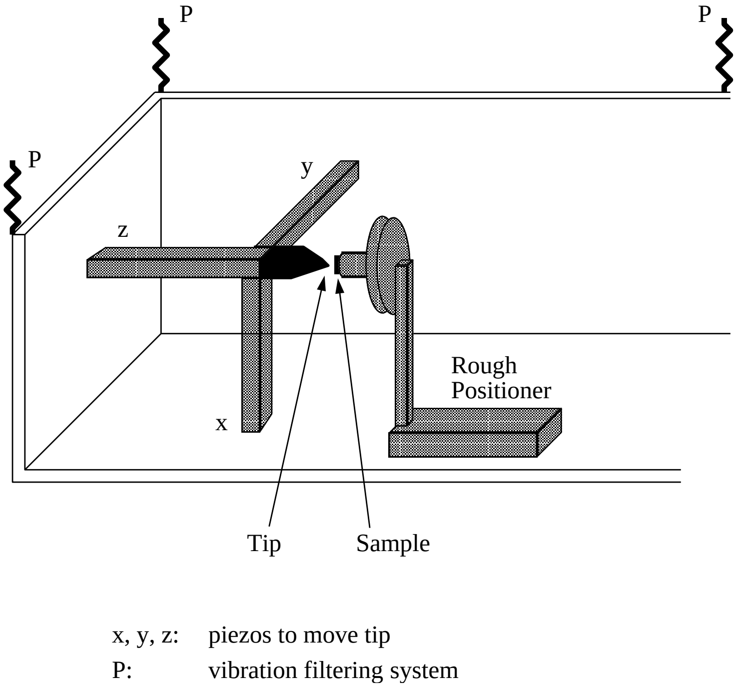


Figure 4.2: A (very) schematic diagram of a scanning tunnelling microscope.

Scanning tunnelling microscopes can measure the interference patterns of electron waves directly. The famous STM image reproduced in Fig. 4.3 shows the electron standing wave inside a quantum “corral” — a circle of 48 Fe atoms arranged using the STM tip itself on the surface of a piece of Cu [M.F. Crommie, C.P. Lutz, D.M. Eigler, *Science* **262**, 218–220 (1993)]. The colours and 3D effect were created from the flat 2D current map using computer graphics, but the wave-like nature of the electron states is clear.

Two-slit interference

Two-slit interference patterns have been created using electrons, atoms and even C_{60} molecules [O. Nairz, M. Arndt, A. Zeilinger, Am. J. Phys. **71**, 319–325 (2003)].

Atomic energy levels

As will be explained Sec. 4.5, the quantised energy levels of atoms are a direct consequence of the wave-like properties of electrons.

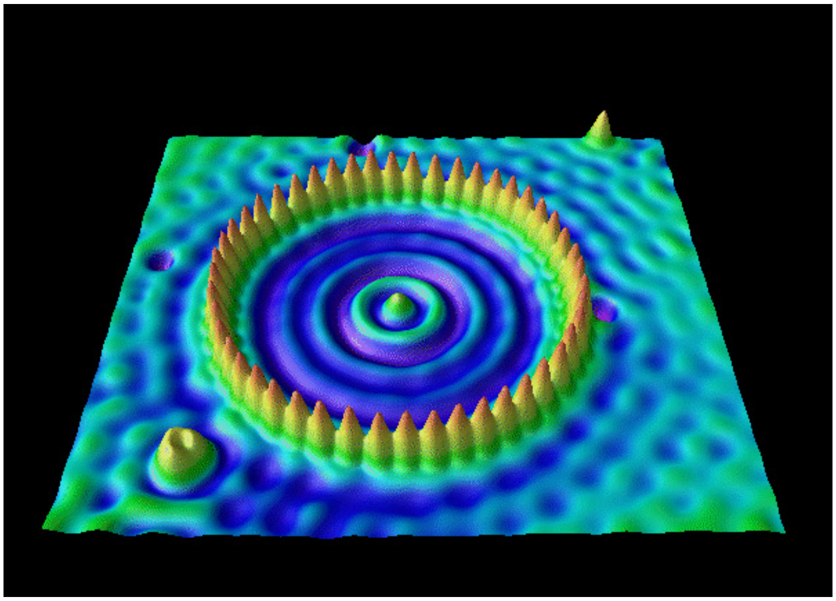


Figure 4.3: A “corral” of 48 Fe atoms on the surface of a piece of Cu. Before the image was taken, the STM tip was used to move the Fe atoms into position one by one. The potential due to the ring of Fe atoms is strong enough to trap an electron within the corral, producing a standing wave pattern.

4.5 Atomic Spectra

As illustrated in Fig. 4.4, hot solids emit/absorb a continuous spectrum of wavelengths/frequencies of electromagnetic radiation. Individual atoms and molecules, by contrast, only emit and absorb at specific wavelengths/frequencies. The sharp absorption/emission lines can be seen by heating a gas, as when a volatile salt is thrown into a flame, or subjecting it to electrical discharges.

The absorption lines of atoms act as fingerprints and can be used to identify elements elsewhere in the universe. Helium, for example, was discovered on Earth only after it had been detected as dark lines in the black-body spectrum of the Sun. The dark lines are created when outgoing black-body photons are absorbed by cool He atoms in the Sun’s outer layers. Atomic absorption lines also played an important role in the detection and measurement of the expansion of the universe: a red-shifted black-body spectrum is just a cooler black-body spectrum and does not tell you the recession velocity of a star unless the real temperature of the star is known; but a red-shifted He or H spectrum is characteristic.

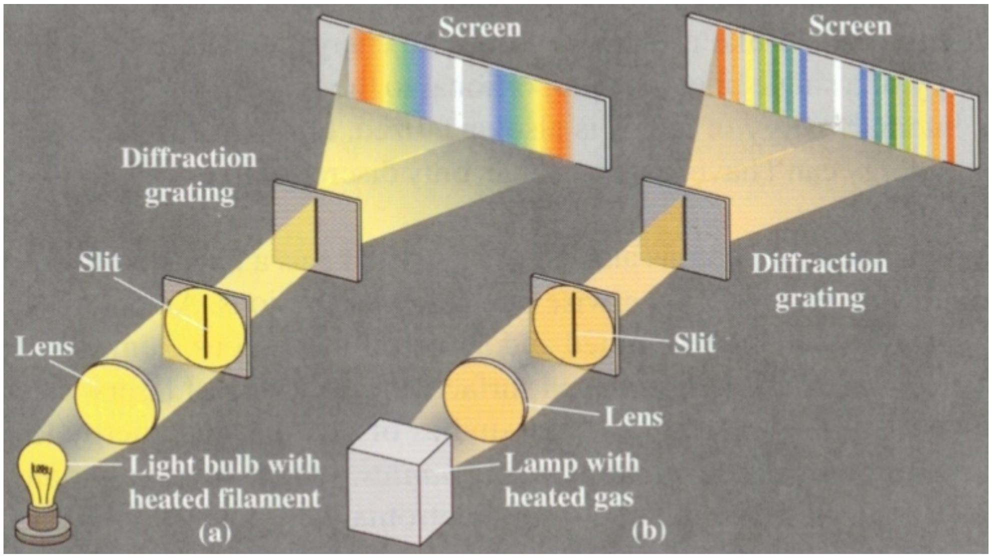


Figure 4.4: Comparison of the absorption spectra of a solid and a gas. [Fig. 38.1 from H.D. Young and R.A. Freedman, *University Physics with Modern Physics*, 11th Ed. (Addison Wesley, 2004).]

The H atom has four clear spectral lines in the visible. These were first observed by Balmer in 1884 and fitted to a simple formula by Rydberg in 1890:

$$\frac{1}{\lambda} = R_{\mathrm{H}}\left(\frac{1}{2^2} - \frac{1}{m^2}\right) , \qquad m = 3, 4, 5, \ldots , \qquad \text{(Balmer series)}$$

where R_{H}, the Rydberg constant, is approximately $1.097 \times 10^7\,\mathrm{m}^{-1}$. Using $1/\lambda = \nu/c$, the Rydberg formula for the Balmer series becomes

$$h\nu = hcR_{\mathrm{H}}\left(\frac{1}{2^2} - \frac{1}{m^2}\right) \approx (13.6\,\mathrm{eV}) \times \left(\frac{1}{2^2} - \frac{1}{m^2}\right) .$$

Replacing $1/2^2$ by $1/1^2$ gives the Lyman series, later observed in the ultraviolet. Replacing $1/2^2$ by $1/3^3$ gives the Paschen series, later observed in the infrared. In fact, spectral lines are found at all frequencies $\nu_{m\to n}$ for which

$$h\nu_{m\to n} \approx (13.6\,\mathrm{eV}) \times \left(\frac{1}{n^2} - \frac{1}{m^2}\right) , \qquad (4.5)$$

with m and n positive integers and $m > n$.

We can explain the spectrum by assuming that an H atom has quantised energy levels,

$$E_n = -\frac{13.6\,\mathrm{eV}}{n^2}\ , \qquad n = 1, 2, \dots\ . \tag{4.6}$$

The energies are all negative because they correspond to bound electronic orbits; the total energies (kinetic + potential) of planets in orbit around the Sun are also negative, for the same reason. The lowest energy level, the *ground state*, has $n = 1$ and energy $-13.6\,\mathrm{eV}$. States with $n > 1$ are known as *excited states*. When an atom makes a transition from a higher energy level m to a lower energy level n, it emits a photon of energy

$$h\nu_{m\to n} \quad = \quad E_m - E_n\ . \tag{4.7}$$

Similarly, when an atom starts in energy level E_n and absorbs a photon of energy $h\nu_{m\to n} = E_m - E_n$, it is excited to energy level E_m.

Other evidence for the quantisation of energy levels is seen in

- X-ray spectra (see Fig. 3.6).
- Vibrational and rotational spectra of molecules. The vibrational energy levels of molecules, such as the CO_2 molecules in the atmosphere, are also quantised and absorb light at specific frequencies only. Microwave ovens excite the rotational energy levels of water molecules.
- Absorption/emission spectra of point defects (such as vacancies, which are missing atoms, or interstitials, which are extra atoms jammed in to the crystal lattice) in solids.

4.6 Why Quantisation?

Quantisation is a natural consequence of confining waves. The harmonics of a violin string make a good example. If the string is bowed normally, the pitch you hear corresponds to the so-called fundamental mode, the wave pattern of which is shown in the top panel of Fig. 4.5. If you touch the string in the middle as you bow, you can excite the first harmonic, shown in the middle panel of Fig. 4.5. The wavelength of the first harmonic is half that of the fundamental, so the frequency is doubled and the note sounds an octave higher. If you touch the string a third of the way along, you can excite the second harmonic, which has a frequency three times higher than the fundamental and sounds an octave and a fifth higher.

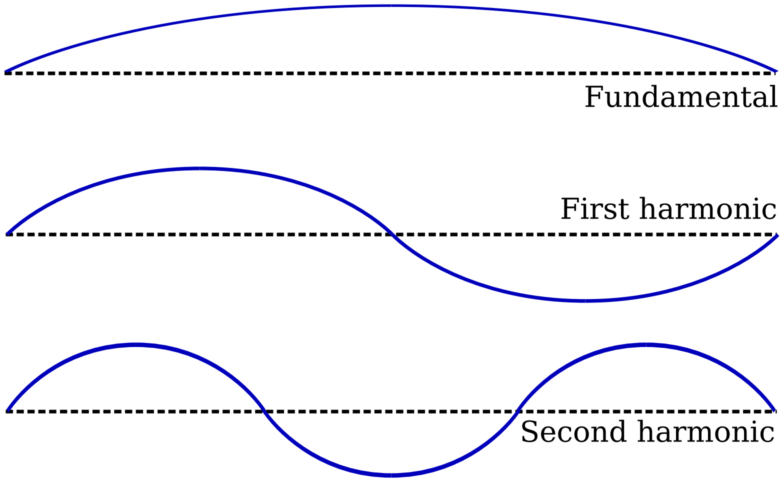


Figure 4.5: The wave patterns of the fundamental, first and second harmonics of a violin string.

The allowed frequencies of a violin string are proportional to n, with $n = 1$ for the fundamental, $n = 2$ for the first harmonic, and so on. This differs from the $1/n^2$ rule that seems to describe the energy levels of an H atom, but the idea is good:

Confinement $\Rightarrow$ only certain standing waves allowed $\Rightarrow$ only certain frequencies allowed $\Rightarrow$ only certain energies allowed ($E = h\nu$).

A solid is a huge box, so the allowed frequencies are very close together and a continuous absorption/emission spectrum is observed.

An atom is a tiny box, so the allowed frequencies are well separated and the absorption/emission spectrum consists of discrete lines.

4.7 The Bohr Atom

The Bohr model was the direct forerunner of modern quantum theory and gives the exact energy levels and rough sizes of one-electron atoms/ions (H, He$^+$, Li^{2+}, Be^{3+}, ...). It does not work for more complicated atoms, molecules or solids, and is probably better regarded as an interesting and useful curiosity than a valid description of nature.

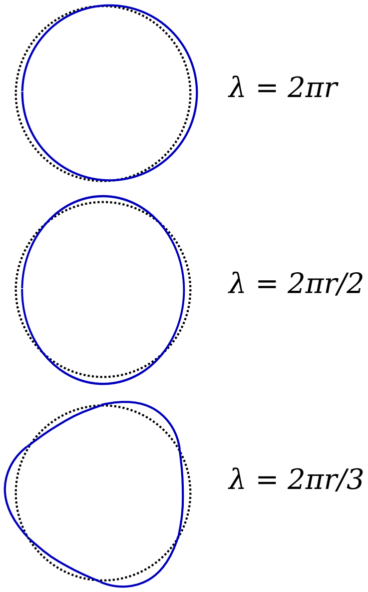


Figure 4.6: The first three Bohr orbits.

The ideas underlying the Bohr model are simple:

(a) The electron moves around the nucleus in a circular Newtonian orbit.

(b) The de Broglie wavelength, $\lambda = h/p$, must "fit in" to the circumference of the orbit.

Consequences of (b)

The second of these ideas is illustrated in Fig. 4.6, which shows that the de Broglie wavelength of the n^{th} Bohr orbit is $2\pi r/n$. Since $p = h/\lambda$, this gives

$$p = \frac{hn}{2\pi r} = \frac{n\hbar}{r} \qquad \text{or} \qquad pr = n\hbar\,.$$

The angular momentum $L = mvr = pr$ is therefore quantised in units of $\hbar$:

$$L \;=\; mvr \;=\; n\hbar\,. \tag{4.8}$$

Angular momentum in real atoms

Although the Bohr theory is wrong about many things, this particular result is more or less right: the angular momentum of an atom is quantised and the unit is $\hbar$. Real atoms are more complicated because the orbits are three-dimensional (the angular momentum $\mathbf{L}$ is actually a vector) and because the electron has a *spin* — an intrinsic "internal" angular momentum of size $\hbar/2$. Furthermore, the quantum mechanical rules for adding the orbital and spin angular momenta differ from the simple vector addition rule used in classical mechanics.

Consequences of (a)

Applying Newton's second law in the form *Force = Centripetal Acceleration* gives

$$\frac{e^2}{4\pi\epsilon_0 r^2} = \frac{mv^2}{r} = \frac{(mvr)^2}{mr^3} \ . \tag{4.9}$$

[Note: this derivation is for a hydrogen atom. The generalisation to a one-electron ion of atomic number Z is covered in Q7 of Problem Sheet 3.] Since $L = mvr = n\hbar$, Eq. (4.9) becomes

$$\frac{e^2}{4\pi\epsilon_0 r_n^2} = \frac{(n\hbar)^2}{mr_n^3} \ ,$$

where r_n is the radius of the n^{th} Bohr orbit. Solving for r_n yields

$$r_n = \frac{4\pi\epsilon_0 (n\hbar)^2}{me^2} \ . \tag{4.10}$$

The smallest orbit has radius $r_1 \approx 0.53 \times 10^{-10}$ m. This distance is called a Bohr radius and denoted a_0. It is about the radius of a small atom.
The energy of the orbiting electron is given by

$$\begin{aligned} E &= \text{KE} + \text{PE} \\ &= \frac{1}{2}mv^2 - \frac{e^2}{4\pi\epsilon_0 r} \\ &= \frac{1}{2}\frac{e^2}{4\pi\epsilon_0 r} - \frac{e^2}{4\pi\epsilon_0 r} \ , \end{aligned}$$

where the last step used Eq. (4.9) again. The energy of the n^{th} Bohr orbit is therefore

$$E_n = \frac{-e^2}{2(4\pi\epsilon_0)r_n} = \frac{-me^4}{2(4\pi\epsilon_0\hbar)^2 n^2} \approx \frac{-(13.6\,\text{eV})}{n^2} \ , \tag{4.11}$$

exactly as the spectroscopic evidence suggests.

Successes and failures of the Bohr Model

The Bohr model

- ✓ provides a partial explanation of the stability of atoms. Orbiting — and thus accelerating — electrons are unable to radiate away energy gradually, as classical physics suggests they should, because they can only have certain quantised energies.
- ✓ gives correct energy levels for the H atom and one-electron ions such as He^+ and Li^{2+}.
- ✓ predicts that angular momentum should be quantised in units of $\hbar$. This is more or less right.
- × is no good for atoms with more than one electron.
- × is no good for molecules or solids.
- × relies on a misleading classical picture of orbiting electrons. The actual behaviour of electrons in atoms is more interesting than the Bohr model suggests.
- × says nothing about how transitions between energy levels occur.

Chapter 5

The Wave Function

5.1 The Two-Slit Experiment Gets Weird

> We choose to examine a phenomenon which is impossible, absolutely impossible, to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery. We cannot make the mystery go away by explaining how it works . . . In telling you how it works we will have told you about the basic peculiarities of all quantum mechanics.
>
> R.P. Feynman, *Lectures on Physics*, Vol. 3, Ch. 1.

To show the strangeness of the quantum world, and to help associate a meaning to the rather abstract wave that seems to be associated with quantum mechanical particles, we are going to analyse three versions of the two-slit experiment:

1. with classical particles
2. with classical waves
3. with quantum mechanical particles

The analysis will show that quantum mechanical “particles” such as electrons and photons are neither particles nor waves, but something wholly new with aspects of both. For want of a better word, they are sometimes called “wavicles”.

In the two-slit experiment, a beam of particles or monochromatic waves passes through a screen containing two long narrow parallel slits. The number density/intensity of particles/waves hitting a detector a long way beyond the screen is measured.

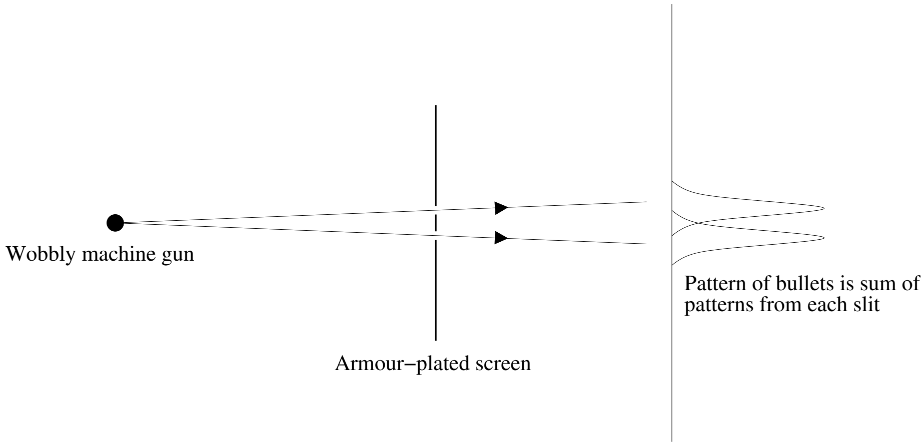


Figure 5.1: The two-list experiment with classical particles.

1. **Classical particles:** Following Feynman, imagine that the incident particles are bullets fired from a wobbly machine gun and that the screen with the slits is armour plated (see Fig. 5.1). The detector consists of a sand trap which catches the bullets and allows us to see where each one lands. After the experiment, the sand trap is found to contain two groups of bullets, one centred on the "shadow" of each slit. (A few bullets may bounce off the sides of the slits and end up outside the shadows.)

   If one slit is closed, the sand trap only collects the group of bullets passing through the open slit. Moreover, the pattern of bullets obtained with both slits open is exactly (to within statistical error) the sum of the patterns obtained with the upper slit closed and with the lower slit closed. If $P_t(x)$ is the probability that a randomly chosen bullet fired from the gun arrives in the detector at x when only the top slit is open, $P_b(x)$ is the probability that it arrives in the detector at x when only the bottom slit is open, and $P_{tb}(x)$ is the probability that it arrives in the detector at x when both slits are open, then

   $$P_{tb}(x) = P_t(x) + P_b(x) \ .$$

2. **Classical waves:** In this case, the machine gun is replaced by a source of waves and the detector by any device that measures the wave intensity (i.e., the arriving energy density) as a function of position. If the classical waves are sound waves, for example, the detector might be an array of microphones. The signal arrives smoothly, not in "lumps" as with classical particles.

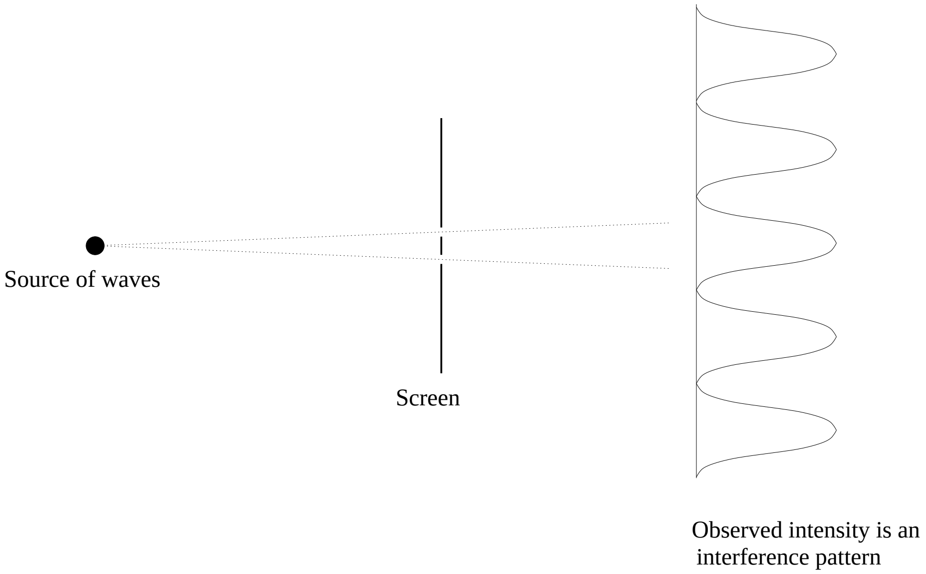


Figure 5.2: The two-slit experiment with classical waves.

The measured intensity is an interference pattern, which may be calculated by adding the complex amplitudes of the waves emerging from the two slits and taking the square modulus of the sum. The form of this interference pattern (in the case where the slits are much narrower than the wavelength) is shown in Fig. 5.2.

The wave theory that leads to this pattern was covered in Sec. 2.3.1. Starting from the observation that the amplitudes of the waves emerging from the two slits add, it was shown that the wave $\psi_{tb}(\zeta, t)$ a distance ζ from the top slit in the θ direction is

$$\begin{aligned}\psi_{tb}(\zeta,t) &= \psi_t(\zeta,t) + \psi_b(\zeta,t) \\ &= Ae^{i(k\zeta-\omega t)} + Ae^{i(k(\zeta+d\sin\theta)-\omega t)} \\ &= A_t e^{i(k\zeta-\omega t)} + A_b e^{i(k\zeta-\omega t)} \ ,\end{aligned}$$

where $A_t = A$ and $A_b = Ae^{ikd\sin\theta}$. Defining $A_{tb} = A_t + A_b$, this becomes

$$\psi_{tb}(\zeta,t) \ = \ A_{tb} e^{i(k\zeta-\omega t)} \ .$$

As usual with waves, the strength of the signal measured by a detector placed far from the slits in the θ direction is proportional to the *intensity*:

$$P_{tb} \ = \ \text{wave intensity in } \theta \text{ direction}$$

$$
\begin{aligned}
&= |A_{tb}|^2 \\
&= (A_t^* + A_b^*)(A_t + A_b) \\
&= |A_t|^2 + |A_b|^2 + A_t^* A_b + A_b^* A_t \\
&= P_t + P_b + (\text{interference terms}).
\end{aligned}
$$

Because the waves from the two slits interfere, the pattern of intensities $P_{tb}(\theta)$ obtained with both slits open is *not* the sum of the pattern $P_t(\theta)$ obtained when only the top slit is open and the pattern $P_b(\theta)$ obtained when only the bottom slit is open.

In fact, if one slit is closed, the interference pattern of the two parallel narrow slits is replaced by the diffraction pattern a single slit. Assuming that the width of the slit is much narrower than the wavelength, the single-slit diffraction pattern is very broad and varies slowly as a function of angle. In places where the intensity of the two-slit interference pattern is zero or very small, closing one slit may therefore *increase* the measured intensity.

3. **Quantum mechanical particles:** In this case, the machine gun is replaced by a source of quantum mechanical particles, perhaps photons from a laser or electrons from an electron gun. The detector is some device (e.g., a CCD camera) able to measure the arrival position of a single photon or electron.

   By stopping down the source until only one quantum particle is in the apparatus at any one time, one can establish that each particle causes a single detector at a single point in space to fire, exactly as for classical particles. However, the distribution of quantum mechanical particles is nothing like the distribution of classical particles; instead, it looks *exactly* like the interference pattern observed for classical waves.

   When one slit is closed, the observed particle distribution changes from the rapidly varying (in space) two-slit interference-pattern to the smooth broad interference pattern of a single narrow slit. As with classical waves, closing one slit causes the wave intensity (which is now proportional to the number of arriving particles) to *increase* at some points. The only way to reconcile this observation with the classical picture of particles as bullets is to allow the paths of the particles passing through one slit to depend on whether the other slit is open or closed. If the paths through one slit are assumed to be independent of the state of the other, closing one slit can only decrease the number of particles hitting the detector at x.

Theories of this correlated type (in which the paths taken through one part of the system depend on what happens far away) can be constructed and made consistent with quantum theory. In the de Broglie-Bohm theory, for example, the paths of the particles are affected by a separate wave, the form of which depends on the state of both slits. Unfortunately, such theories are highly non-local: you cannot in general predict the path of a particle without knowing the position of more or less every other particle in the universe. The wave interference picture used in standard quantum theory is simpler and more physical.

Conclusions

The quantum mechanical two-slit experiment suggests that quantum mechanical objects are associated with waves, the amplitudes of which add linearly. Just as in classical physics, the superposition of the waves emerging from the two slits yields an interference pattern.

The two-slit experiment also shows that quantum mechanical objects behave like particles. The detectors that measure the arrival positions of these quantum particles always find them at one place or another — their energy is localised, not smeared out as for a wave.

The link between the particle and wave pictures is provided by the wave intensity: the probability that a position measurement finds a particle at point $\mathbf{r}$ is proportional to the wave intensity at that point. This is typical of quantum theory: although it provides precise predictions of probabilities, it cannot predict where a particular particle will end up.

The interference pattern of arrival probabilities does not depend on the number of particles passing through the apparatus per second. If the particle flux is very low, very few dots appear on the screen per second, but the distribution of dots is unaltered. The interference pattern persists even when the flux is so low that there is no significant chance of finding more than one particle in the apparatus at a time. This shows that the interference cannot be the result of many particles "colluding" — one is all you need.

5.2 What is an Electron?

An electron is . . .

- *not* a particle (although it has some particle-like properties);

- *not* a wave (although it has some wave-like properties).

It is something new — a “wavicle”.

- In between measurements, wavicles propagate like waves, diffracting, refracting, interfering, and so on. The Schrödinger equation that describes how these waves propagate is not quite the same as the classical wave equation, but the two are closely related.
- Some types of detectors respond to wavicles by producing particle-like “spikes” at sharply defined positions. After such a measurement, the wave packet of the wavicle is very small (but it soon begins to spread out again — as wave packets do).

Questions quantum theory cannot answer:

- Where is the particle between measurements?
- How did the particle get from A to B?

Probabilities in quantum theory:

- Quantum mechanics allows you to work out the possible values (known as *eigenvalues*) of any observable (measurable quantity). A few important observables are position, momentum, velocity, and energy.
- It you know the wave function $\psi(x,t)$, you can also work out the probabilities of measuring each of these allowed values.
- The wave function $\psi(x,t)$ is central to quantum theory but is not itself an observable. The intensity $|\psi(x,t)|^2$ is the probability density observed in position measurements.

The phenomenon of radioactive decay provides some evidence in favour of the idea that the world may really be probabilistic at the atomic scale. Suppose, for example, that you start with a large number $N(t{=}0)$ of radioactive nuclei. Instead of all decaying together a fixed time after they were created, as might be expected, the number $N(t)$ remaining at time t decreases exponentially: $N(t) = N(0)e^{-\alpha t}$, where α is a constant. This observation can be explained if one assumes that the nuclei decay at random, with a fixed fraction α decaying per second:

$$\begin{aligned} dN &= -\alpha N dt \\ \Rightarrow \qquad \int \frac{dN}{N} &= -\alpha \int dt \\ \Rightarrow \qquad \ln N(t) &= C - \alpha t\ . \end{aligned}$$

Setting $t = 0$ shows that $C = \ln(N(0))$ and hence

$$N(t) = N(0)e^{-\alpha t}\ .$$

5.3 Review of Probability Theory

Because quantum mechanics predicts only probabilities, concepts from probability theory play a central role in the rest of this course. Most first years find these concepts straightforward, but past experience has shown that some do not; the idea of a probability density, in particular, causes widespread confusion. To help those of you who have not encountered probability densities before, this section provides a brief, self-contained, review of the aspects of probability theory required for this course.

5.3.1 Probabilities

Start by thinking about a game of darts. Suppose that you throw N darts at a dart board and record the scores. The results of the first 21 throws are shown in Fig. 5.3.

Another way to summarise the data is to plot the number n_s of darts scoring each possible value s. An example is shown in Fig. 5.4. Since no single dart can score more than 60, $n_s = 0$ for all $s > 60$.

The total number of darts thrown is equal to the number n_0 that scored 0 plus the number n_1 that scored 1 plus the number n_2 that scored 2 plus . . .

$$n_0 + n_1 + n_2 + \ldots = N\ ,$$

or, equivalently,

$$\sum_{\text{score } s=0}^{\infty} \frac{n_s}{N} = 1\ .$$

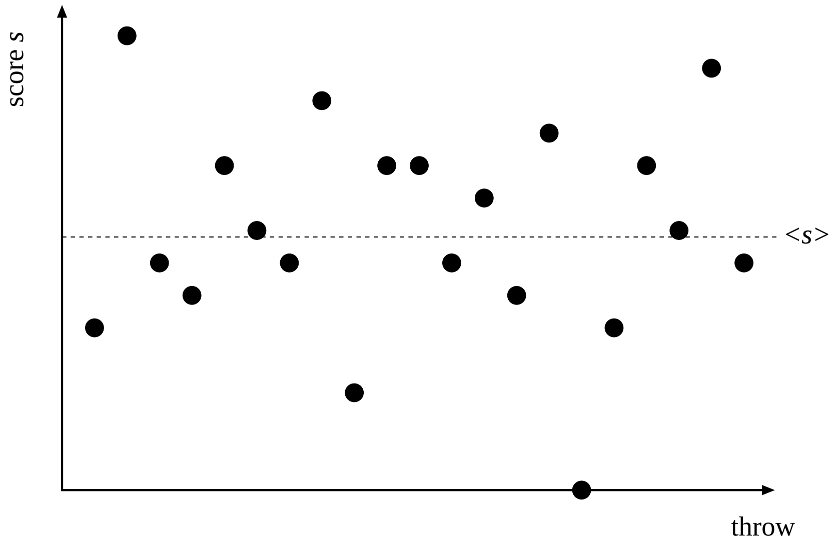


Figure 5.3: The first 21 scores recorded in the dart game

In the limit of large N (a very long game), the ratio n_s/N normally becomes independent of N (assuming that the player does not get tired!). This limiting ratio is the called the *probability*, p_s, of scoring s:

$$p_s \ := \ \lim_{N\to\infty} \frac{n_s}{N} \ .$$

It follows that

$$\sum_{s=0}^{\infty} p_s \ = \ 1 \ .$$

5.3.2 Expected values

The mean or expected score per dart, denoted $\langle s \rangle$, is given by

$$\langle s \rangle \ = \ \lim_{N\to\infty} \left(\frac{1}{N} \times \text{(sum of scores of all } N \text{ darts)} \right) \ .$$

Since n_0 darts scored 0, n_1 scored 1, n_2 scored 2, and so on, we have

$$\text{sum of scores of all } N \text{ darts} \ = \ \sum_{s=0}^{\infty} s n_s \ .$$

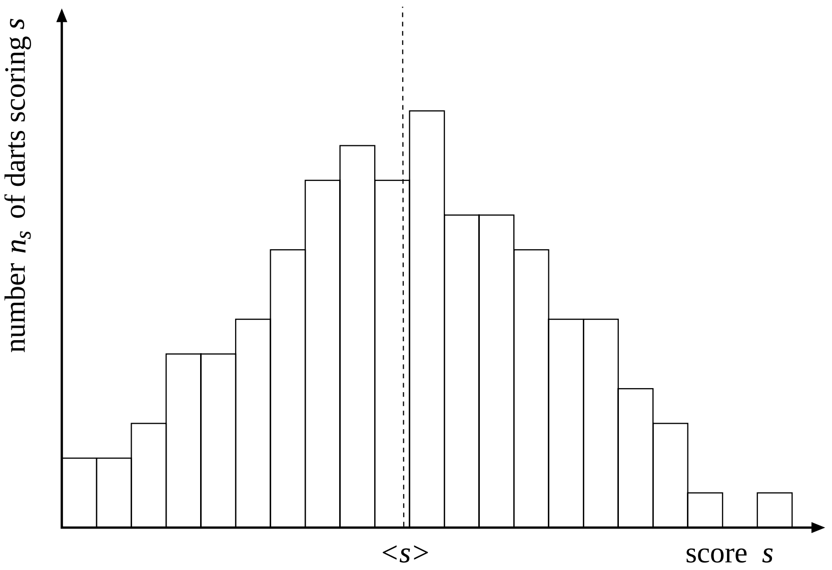


Figure 5.4: The number of darts n_s scoring each possible value of s

This enables us to re-express $\langle s\rangle$ in terms of probabilities:

$$\langle s\rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{s=0}^{\infty} s n_s = \sum_{s=0}^{\infty} s p_s \ .$$

In other words,

$$\begin{aligned}\langle s\rangle &= (0\times \text{Prob. of scoring } 0) + (1\times \text{Prob. of scoring } 1) + \ldots \\ &\quad + (s\times \text{Prob. of scoring } s) + \ldots \ .\end{aligned}$$

More generally, the mean or expected value of any function $g(s)$ of the score s is given by:

$$\langle g(s)\rangle = \lim_{N\to\infty} \frac{\sum_{s=0}^{\infty} g(s) n_s}{N} = \sum_{s=0}^{\infty} g(s) p_s \ .$$

For example, the expected value of s^2,

$$\langle s^2\rangle = \lim_{N\to\infty} \frac{\sum_{s=0}^{\infty} s^2 n_s}{N} = \sum_{s=0}^{\infty} s^2 p_s \ .$$

5.3.3 Variance and standard deviation

Evaluating $\langle s\rangle$ tells us the mean score per dart but provides no information about the spread of scores. We would also like to know the typical distance of a point in Fig. 5.3 from the horizontal line $s = \langle s\rangle$ and the typical width of the distribution in Fig. 5.4.

The simplest possible measure of spread or uncertainty, the expected value of $s - \langle s\rangle$, is no good:

$$\begin{aligned}
\langle s - \langle s\rangle\rangle &= \sum_{s=0}^{\infty}(s - \langle s\rangle)p_s \\
&= \sum_{s=0}^{\infty} sp_s - \langle s\rangle \sum_{s=0}^{\infty} p_s \\
&= \langle s\rangle - \langle s\rangle \quad \text{(remember that } \textstyle\sum_{s=0}^{\infty} p_s = 1) \\
&= 0\ .
\end{aligned}$$

Because of the way $\langle s\rangle$ is defined, the positive and negative contributions cancel and $\langle s - \langle s\rangle\rangle$ is zero.

One way to avoid this cancellation would be to work out the expected value of $|s - \langle s\rangle|$, but the modulus function is mathematically awkward because the slope of $|x|$ changes discontinuously as x passes through the origin. A more convenient measure of the width of a probability distribution is the *standard deviation* σ, defined to be the (positive) square root of the *variance*

$$\sigma^2 = \left\langle (s - \langle s\rangle)^2 \right\rangle .$$

In words: the variance σ^2 is the expected value of the square of the distance of the score s from the mean score $\langle s\rangle$; and the standard deviation σ is the square root of the variance. The standard deviation is also called the root-mean-square (rms) width of the probability distribution. Unlike the variance, the standard deviation always has the same physical dimensions as the random variable s. (Both σ and σ^2 are dimensionless in our darts example.)

Another useful formula for the variance is

$$\sigma^2 = \langle s^2\rangle - \langle s\rangle^2 .$$

This can be derived from the original definition as follows:

$$\left\langle (s - \langle s\rangle)^2 \right\rangle = \sum_{s=0}^{\infty} (s - \langle s\rangle)^2 p_s$$

$$
\begin{aligned}
&= \sum_{s=0}^{\infty} \left(s^2 - 2\langle s\rangle s + \langle s\rangle^2\right) p_s \\
&= \sum_{s=0}^{\infty} s^2 p_s - 2\langle s\rangle \sum_{s=0}^{\infty} s p_s + \langle s\rangle^2 \sum_{s=0}^{\infty} p_s \\
&= \langle s^2\rangle - 2\langle s\rangle\langle s\rangle + \langle s\rangle^2 \\
&= \langle s^2\rangle - \langle s\rangle^2 \; .
\end{aligned}
$$

5.3.4 Probability densities

Until now we have been considering quantities such as scores in a dart game that can only take discrete (separate, quantised) values. A few small adjustments are required to apply the ideas of probability theory to continuous variables such as the heights of people.

Instead of throwing N darts, imagine that you measure the heights of N people. The probability that anybody in your sample is exactly 1.8 m tall is zero. Some people may be roughly 1.8 m tall, and a few may be very close to 1.8 m tall, but there is no chance of finding someone who is *exactly* 1.8 m tall (plus or minus nothing).

We can, however, ask about the number of people $n(h, h + \Delta h)$ with heights between h and $h + \Delta h$, where Δh is finite. Given a large enough sample of people, the ratio $n(h, h + \Delta h)/N$ is independent of the sample size N and we can define the corresponding probability

$$p(h, h + \Delta h) \; = \; \lim_{N\to\infty} \frac{n(h, h + \Delta h)}{N} \; ,$$

just as in the discrete case.

If Δh is small enough, the number of people with heights between h and $h + \Delta h$ ought to be proportional to Δh. (For example, one would expect the number of people with heights between 1.8000 m and 1.8002 m to be roughly twice the number with heights between 1.8000 m and 1.8001 m.) This suggests defining a *probability density function* (pdf) $f(h)$ via:

$$p(h, h + \Delta h) \; = \; f(h)\,\Delta h \; .$$

As long as Δh is small enough, $f(h)$ should be independent of the value of Δh. The pdf is the central quantity in all applications of probability theory to continuous random variables.

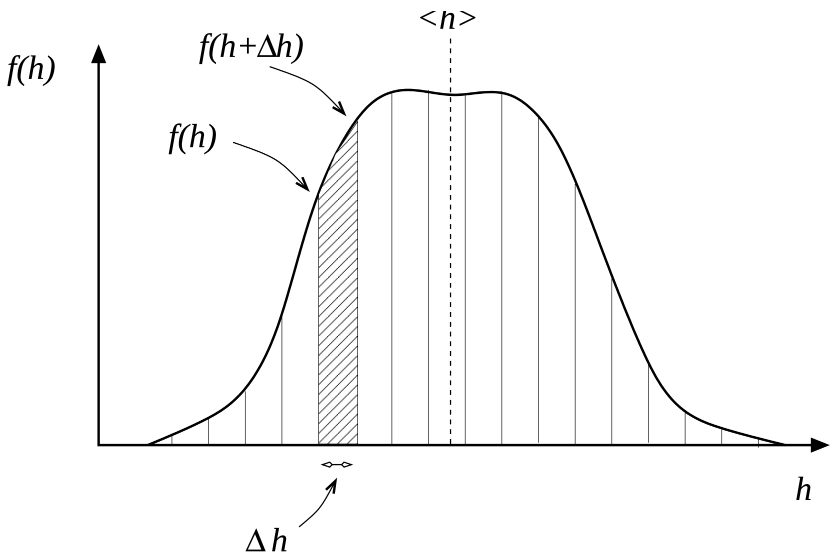


Figure 5.5: A probability density function

Figure 5.5 shows an example pdf. The probability that the height of a randomly chosen person lies between h and $h + \Delta h$ is equal to the shaded area, which is approximately $f(h)\Delta h$ if Δh is small enough. The probability that the height lies between $h_{\rm low}$ and $h_{\rm high}$ (where $h_{\rm low}$ and $h_{\rm high}$ need not be close) is the area under the pdf from $h_{\rm low}$ to $h_{\rm high}$:

$$p(h_{\rm low}, h_{\rm high}) \;=\; \int_{h_{\rm low}}^{h_{\rm high}} f(h)\, dh \;.$$

Since the probability that h lies somewhere between zero and infinity is equal to 1, it follows that

$$\int_0^\infty f(h)\, dh \;=\; 1 \;.$$

Following the darts example, the expected height $\langle h \rangle$ is defined via:

$$\begin{aligned} \langle h \rangle \;\approx\; & (0 \times \text{Prob. height is between 0 and } \Delta h) \\ & + (\Delta h \times \text{Prob. height is between } \Delta h \text{ and } 2\Delta h) \\ & + \;\ldots \\ & + (n\Delta h \times \text{Prob. height is between } n\Delta h \text{ and } (n+1)\Delta h) \end{aligned}$$

$$
\begin{aligned}
&+ \dots \\
\approx\ & \sum_{n=0}^{\infty} n\Delta h\, f(n\Delta h)\, \Delta h \\
=\ & \sum_{n=0}^{\infty} h_n f(h_n)\, \Delta h \ ,
\end{aligned}
$$

where $h_n := n\Delta h$ is the value of the height h at the left-hand edge of the n^{th} strip of width Δh. In the limit as $\Delta h \to 0$, the summation turn into an integral and the $\approx$ signs become $=$ signs:

$$
\langle h \rangle \ = \ \int_0^\infty h f(h)\, dh \ .
$$

Just as in the discrete case, the expected value of any function $g(h)$ of the height h is given by:

$$
\langle g(h) \rangle \ = \ \int_0^\infty g(h) f(h)\, dh \ .
$$

The variance σ^2 is defined exactly as in the discrete case:

$$
\sigma^2 \ = \ \left\langle (h - \langle h \rangle)^2 \right\rangle \ = \ \langle h^2 \rangle - \langle h \rangle^2 \ ,
$$

but the expectation values are now integrals,

$$
\sigma^2 \ = \ \int_0^\infty (h - \langle h \rangle)^2\, f(h)\, dh \ = \ \int_0^\infty h^2\, f(h)\, dh - \left(\int_0^\infty h\, f(h)\, dh \right)^2 \ ,
$$

instead of summations. The standard deviation is still the square root of the variance.

5.4 Working with Wave Functions

Suppose that a particle (wavicle) has wave function $\psi(x)$ at some fixed time t (suppressed to simplify the notation). Our analysis of the two-slit experiment suggested that

$$
|\psi(x)|^2 dx \ \ \propto \ \ \begin{cases} \text{probability that a measurement of position gives} \\ \text{a result between } x \text{ and } x + dx \end{cases}
$$

A wave function is said to be *normalised* if

$$\int_{-\infty}^{\infty} |\psi(x)|^2 dx \;=\; 1 \;. \tag{5.1}$$

If $\psi(x)$ is normalised, the "$\propto$" sign can be replaced by an "$=$" sign,

$$|\psi(x)|^2 dx \;=\; \begin{cases} \text{probability that a measurement of position gives} \\ \text{a result between } x \text{ and } x + dx \end{cases} \tag{5.2}$$

and $|\psi(x)|^2$ is a probability density.

Given an unnormalised wave function $\psi_u(x)$, for which

$$\int_{-\infty}^{\infty} |\psi_u(x)|^2 dx \;=\; C \qquad (C \neq 1) \;,$$

we can find the corresponding normalised wave function $\psi_n(x)$ by dividing through by $\sqrt{C}$:

$$\psi_n(x) = \frac{\psi_u(x)}{\sqrt{C}} \;.$$

It is usually easiest to work with normalised wave functions.

Most of the examples in this course assume that the world is one-dimensional, but three-dimensional wave functions are only a little more difficult to deal with. A three-dimensional wavefunction $\psi(\mathbf{r})$ is normalised if

$$\int_{\text{all space}} |\psi(\mathbf{r})|^2 \, d^3r \;=\; 1 \;.$$

As in one dimension, if $\psi(\mathbf{r})$ is normalised then $|\psi(\mathbf{r})|^2$ is a probability density:

$$|\psi(\mathbf{r})|^2 \, d^3r \;=\; \begin{cases} \text{probability that a measurement of position gives} \\ \text{a result in the volume element } d^3r \text{ at position } \mathbf{r}. \end{cases}$$

Example: Particle in a Box

A particle in the box-like potential well

$$V(x) \;=\; \begin{cases} 0 & 0 < x < a, \\ \infty & \text{otherwise.} \end{cases}$$

has the ground-state wave function

$$\psi_u(x) \;=\; \begin{cases} \sin\left(\pi x/a\right) & 0 < x < a, \\ 0 & \text{otherwise.} \end{cases}$$

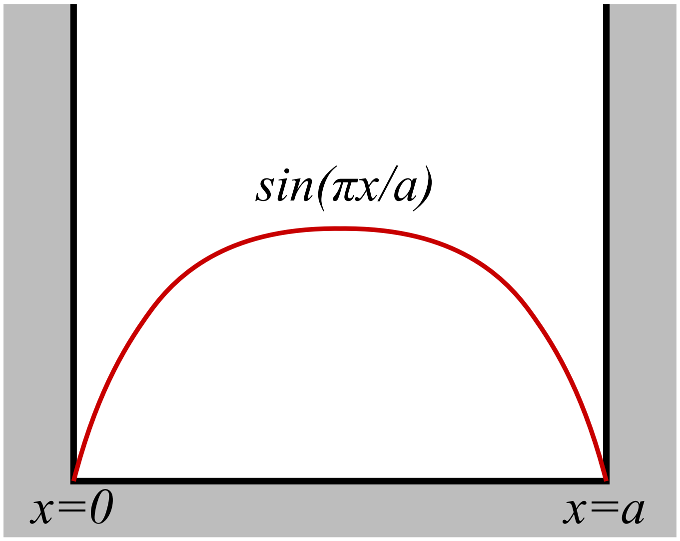


Figure 5.6: The unnormalised ground-state wave function of a particle confined in a box.

(a) Normalise ψ_u:

$$\int_{-\infty}^{\infty} |\psi_u(x)|^2 dx \;=\; \int_0^a \sin^2\left(\frac{\pi x}{a}\right) dx \;=\; \frac{1}{2}a\,,$$

so the normalised wave function is

$$\psi_n(x) \;=\; \begin{cases} \sqrt{\frac{2}{a}}\sin\left(\frac{\pi x}{a}\right) & 0 < x < a, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Find the mean position of the particle:

$$\begin{aligned} \langle x \rangle &= \int_{-\infty}^{\infty} x\,|\psi(x)|^2\,dx \\ &= \int_0^a x\,\frac{2}{a}\sin^2\left(\frac{\pi x}{a}\right)\,dx \\ &= \text{lots of algebra (try it!)} \\ &= \frac{a}{2}\,. \end{aligned}$$

This makes sense.

(c) Find the rms uncertainty in the position of the particle:

$$\begin{aligned}
(\Delta x)^2 &= \sigma^2 = \left\langle (x - \langle x \rangle)^2 \right\rangle \\
&= \int_0^a (x - \langle x \rangle)^2 \, \frac{2}{a} \sin^2\left(\frac{\pi x}{a}\right) \, dx \\
&= \int_0^a \left(x - \frac{a}{2}\right)^2 \frac{2}{a} \sin^2\left(\frac{\pi x}{a}\right) \, dx \\
&= \text{even more algebra (try it again!)} \\
&= a^2 \left(\frac{1}{12} - \frac{1}{2\pi^2}\right) \, .
\end{aligned}$$

Hence

$$\Delta x \approx 0.18a \, .$$

Chapter 6

Wave Packets and the Uncertainty Principle

6.1 Fourier Decomposition

A wave packet is any group of waves. It does not have to be neat and symmetrical and centred on the origin — all that matters is that it dies away to zero far from some centre. Because a QM particle is represented by a wave packet of de Broglie waves, wave packets play an important role in quantum physics.

The most interesting wave packets have a clearly distinguishable "carrier" wave, the amplitude of which is modulated by a much more slowly varying envelope. The wave packet in Fig. 6.1 is of this type. If it were a sound, the pitch would be the frequency of the carrier wave and the square of the envelope would give the volume as a function of time.

Not all wave packets are so simple. Figure 6.2 shows a messier one without a clear carrier wave. A sound of this type would be a noise — a hand clap or a door closing — rather than a musical note with a clear pitch.

Wave packets are not simple cosine waves, but it is plausible (and true) that any function $\psi(x)$ can be expressed as a sum of cosines with different wave vectors, amplitudes and phases:

$$\psi(x) = \sum_{n=1}^{N} a_n \cos(k_n x + \phi_n) = \mathrm{Re}\left(\sum_{n=1}^{N} a_n e^{i(k_n x + \phi_n)}\right) = \mathrm{Re}\left(\sum_{n=1}^{N} A_n e^{ik_n x}\right) ,$$

where $A_n := a_n e^{i\phi_n}$. By choosing the wave vectors, amplitudes and phases carefully, it is possible to ensure that the waves in the superposition undergo perfect

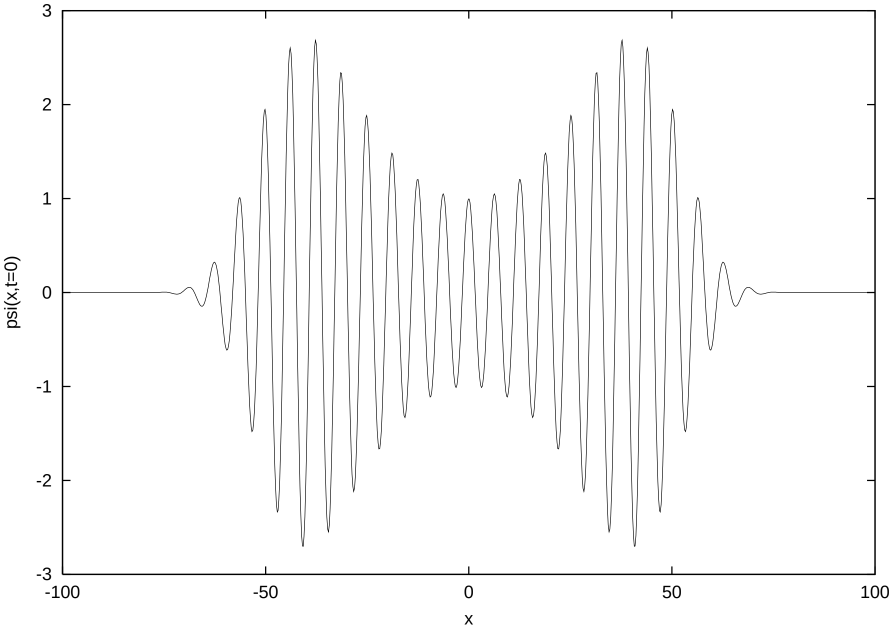


Figure 6.1: A wave packet with a clear carrier wave

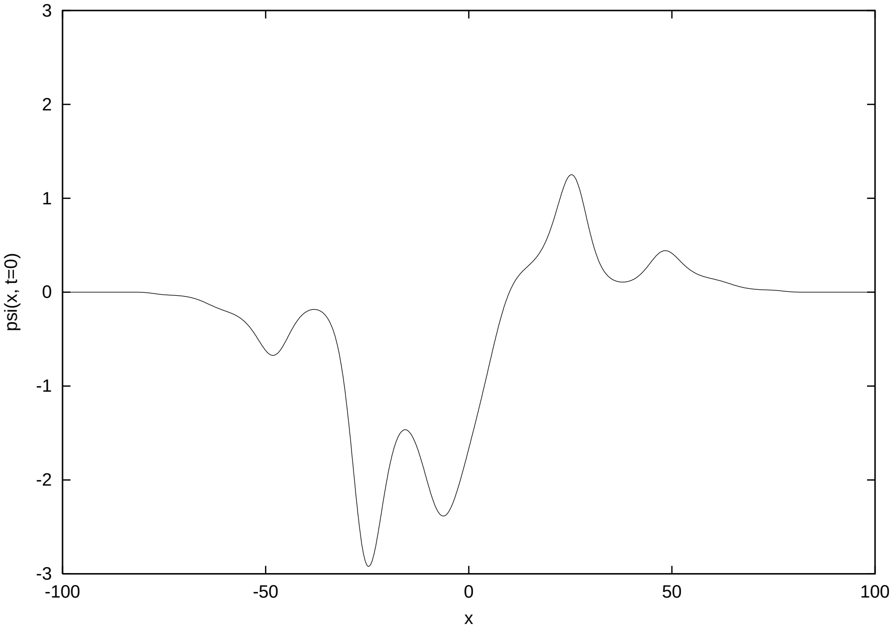


Figure 6.2: A wave packet without a clear carrier wave

destructive interference (producing a total amplitude of zero) everywhere except in one small region of space. The function $\psi(x)$ is then a wave packet.

Actually, as you might already have guessed, the description above is slightly misleading: it is not in fact possible to construct a wave packet using a finite number of cosine waves. To ensure perfect destructive interference *everywhere* far away from the centre, it is necessary to superpose cosine waves of all possible wave vectors k. The sum over wave vectors then becomes an integral and the wave packet takes the form

$$\psi(x) \;=\; \mathrm{Re}\left(\int_{-\infty}^{\infty} a(k)\, e^{i(kx+\phi(k))}\, dk\right) \;=\; \mathrm{Re}\left(\int_{-\infty}^{\infty} A(k)\, e^{ikx}\, dk\right)\,, \qquad (6.1)$$

where $a(k)$ is the real amplitude and $A(k) := a(k)e^{i\phi(k)}$ is the complex amplitude of the component of wave vector k. The fact that any function $\psi(x)$ can be expressed as an integral of cosine waves is plausible, I hope, but the prospect of having to work out the amplitudes $a(k)$ and phase shifts $\phi(k)$ may appear forbidding. In fact, as you will learn when you study Fourier analysis at the beginning of next year, it turns out to be quite easy. (Incidentally, you already know the mathematics needed to understand Fourier analysis, so there is nothing to stop you looking it up in a book if you are interested.) For the purposes of this course, all you need to know is that any function can be obtained by superposing enough cosine waves.

Now that the idea of Fourier superposition has been introduced, the difference between the wave packets in Figs. 6.1 and 6.2 can be described in mathematical terms. In Fig. 6.1 (the musical instrument), all the wavelengths $\lambda = 2\pi/k$ appearing in the superposition are very close to the carrier wavelength. This explains why the ear is able to pick out the carrier frequency and associate a pitch with the sound. In Fig. 6.2 (the hand clap), the wavelengths appearing in the Fourier superposition are all over the place. Since they are no longer clustered around a central carrier wavelength, the ear cannot pick out a clear pitch.

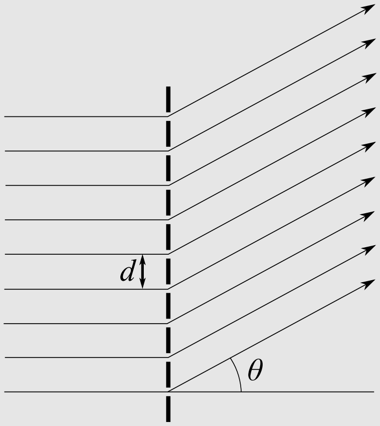
This observation has some interesting repercussions. Imagine, for example, that you want to make a wave packet that "sounds like" middle C, the frequency of which is approximately 262 Hz. If the wave packet is to have a clearly distinguishable pitch, it has to be long enough to contain many carrier-wave oscillations of this frequency. The exact number depends on how cleverly the human brain interprets sounds, but 25 might be a reasonable guess. Such a wave packet takes about $25/262\,\mathrm{s}$ to pass by. Hence, no noise significantly shorter than 0.1 seconds can possibly sound like middle C. This is why most percussion instruments, which make very short sounds, have no discernible pitch.

Physical meaning of $A(k)$

The individual waves (often known as "components") in

$$\psi(x) = \sum_n A_n e^{ik_n x}$$

may be separated using a diffraction grating.



The component of wave vector k_n produces sharp beams where

$$d \sin\theta = m\lambda_n \ , \qquad m = 0, 1, \ldots \qquad (\lambda_n = 2\pi/k_n).$$

In essence, the grating is a momentum-measuring device. Unlike the incident particles, which were described by wave packets and therefore had an uncertain momentum, an outgoing particle found in a particular diffracted beam must have wavelength λ_n and momentum $p_n = h/\lambda_n = \hbar k_n$. Since the intensity of the λ_n beams is proportional to $|A_n|^2$, we conclude that

> The probability that a measurement of p yields the result $\hbar k_n$ is proportional to $|A_n|^2$.

This suggests that $A(k)$ is like a "wave function for momentum".

6.2 The Bandwidth Theorem

The observation that short wave packets do not have clear frequencies is not just an aspect of human perception; it is a mathematical fact. If the interference of the contributions from different wave vectors k to the integral in Eq. (6.1) is to change from constructive at the centre of the wave packet to destructive only a short distance away, the range of wave vectors k (and thus angular frequencies ω) must be wide.

Two-wave example

Let us investigate this idea mathematically. Suppose that a wave packet is constructed using cosine waves with wave vectors in a narrow range $k_c \pm \Delta k$ centred on the carrier wave vector k_c. What is the minimum possible size of such a wave packet? For simplicity, start by considering a simple example with only two component cosine waves, $\psi_1(x)$ and $\psi_2(x)$, which have equal amplitudes and interfere constructively at $x = 0$:

$$\psi_1(x) \;=\; \cos(k_1 x)\;, \qquad \psi_2(x) \;=\; \cos(k_2 x)\;.$$

Figure 6.3 shows the two components and their sum,

$$\psi(x) \;=\; \psi_1(x) + \psi_2(x) \;=\; \cos(k_1 x) + \cos(k_2 x)\;, \tag{6.2}$$

when $k_1 = 0.95\,\mathrm{m}^{-1}$ and $k_2 = 1.05\,\mathrm{m}^{-1}$. The short carrier wave and the slowly varying interference envelope of the beats are clear.

How wide are the beats? The first zero of the envelope function occurs at the point marked by the dashed vertical line on Fig. 6.3. The functions $\cos(k_1 x)$ and $\cos(k_2 x)$ are equal and opposite at this point, and hence the phases $k_1 x$ and $k_2 x$ differ by π. This gives

$$k_2 x = k_1 x + \pi\;,$$

or, equivalently,

$$\text{half-width of beat envelope} \;=\; \frac{\pi}{|k_2 - k_1|}\;.$$

The smaller the difference $\Delta k = |k_2 - k_1|$ between the wave vectors of the two cosine waves, the larger the half-width Δx of the wave packet:

$$\Delta x \Delta k = \pi\;.$$

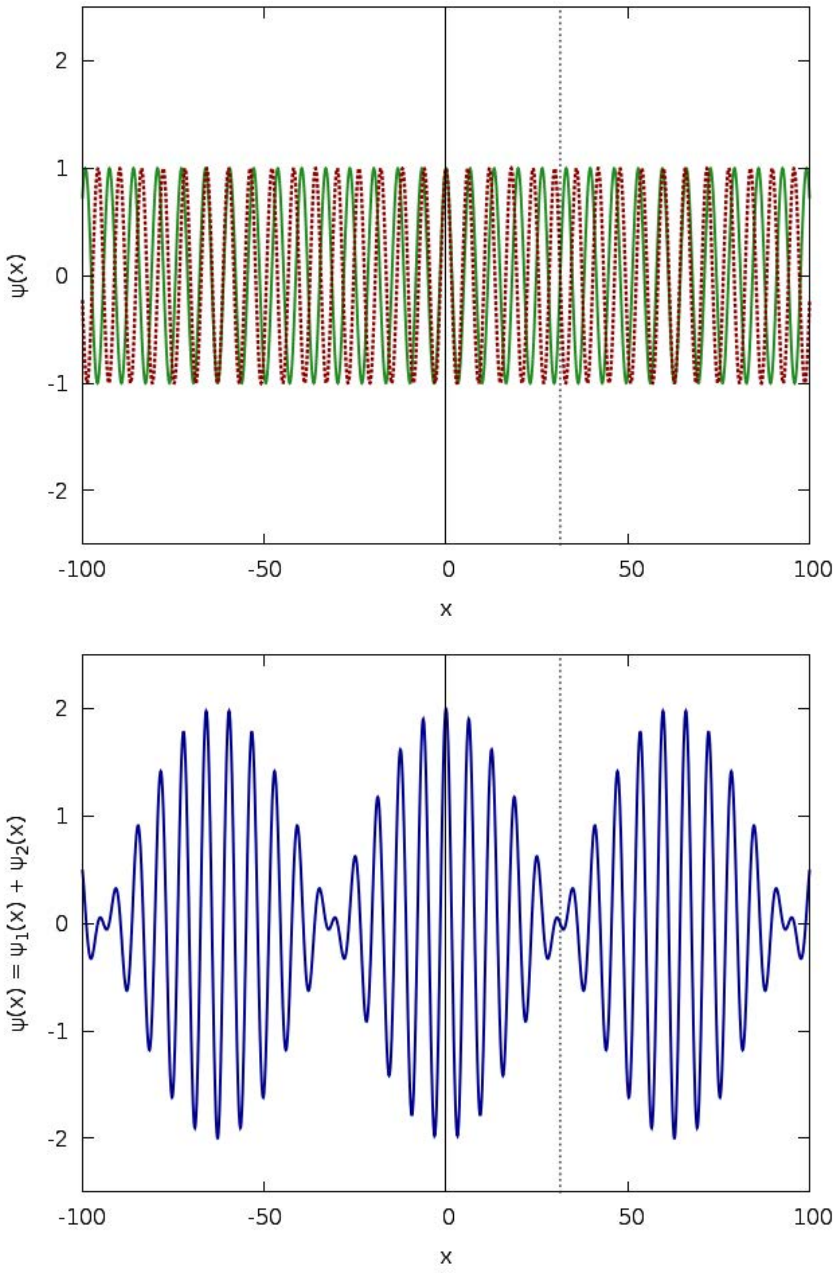


Figure 6.3: Two cosine waves with similar wavelengths and their interference pattern. The vertical dashed lines mark the point where $1.05x = 0.95x + \pi$.

General case

Let us now consider a general wave packet in the form of Eq. (6.1),

$$\psi(x) \;=\; \int A(k)e^{ikx}\,dk \;, \tag{6.3}$$

constructed using an $A(k)$ peaked around some central "carrier" wave vector k_c. (We have omitted the "Re" symbol from Eq. (6.3) for simplicity.) The mathematics is harder in this case, but the physical principle is identical: narrowing the spread Δk of wave vectors contributing to the wave packet necessarily increases its width Δx.

Given the interpretation of $A(k)$ as a wave function for momentum, it makes sense to normalise it such that

$$\int |A(k)|^2\,dk \;=\; 1 \;.$$

We can then define k_c as the mean value of k,

$$k_c \;=\; \langle k\rangle \;=\; \int k\,|A(k)|^2\,dk \;,$$

and Δk as its rms spread,

$$(\Delta k)^2 \;=\; \int (k-\langle k\rangle)^2\;|A(k)|^2\,dk \;.$$

In order to shift from the constructive interference that occurs at the centre of the wave packet to the perfect destructive interference that occurs everywhere outside the wave packet, the relative phases of the components have to change by something close to π. This allows us to estimate the rms width of the wave packet, Δx, from the rms spread of wave vectors, Δk. The distance Δx over which the components with wave vectors $k_c-\Delta k$ and $k_c+\Delta k$ accumulate a phase difference of π satisfies

$$(k_c+\Delta k)\Delta x \;=\; (k_c-\Delta k)\Delta x + \pi \;,$$

and hence

$$\Delta x \;\approx\; \frac{\pi}{2\Delta k} \;.$$

The rms width of the wave packet cannot be very much smaller than this, no matter how cleverly the phases and amplitudes are chosen.

This reasoning is imprecise (you might quibble about the definition of Δk or about factors of 2 or π), but the conclusion is simple and correct. In fact, if $\psi(x)$ is normalised and $\langle x\rangle$ and Δx are defined via

$$\begin{aligned} \langle x\rangle &= \int x\,|\psi(x)|^2\,dx\ , \\ (\Delta x)^2 &= \int (x-\langle x\rangle)^2\ |\psi(x)|^2\,dx\ , \end{aligned}$$

the mathematics of Fourier analysis proves that the values of Δx and Δk are related by a strict inequality,

$$\boxed{\Delta x\,\Delta k \ \geq\ \frac{1}{2}} \tag{6.4}$$

This is known as the bandwidth theorem. It is important to understand that no quantum mechanics is involved here: the bandwidth theorem is a general result and applies to all sorts of waves.

Note that nothing has been said about the *maximum* size of the wave packet. It is always possible to choose the phases and amplitudes of the components such that they interfere constructively at any number of arbitrary places, so the maximum size is unbounded.

Real wave packets move around and change shape, so they are functions of x and t, not just x. However, given a time-dependent wave packet $\psi(x,t)$, you can always create a snapshot $\psi(x)$ by choosing a time t_0 and defining $\psi(x) := \psi(x,t_0)$. The shape of $\psi(x)$ depends on the choice of t_0, but the bandwidth theorem holds for *any* function of x and thus for any snapshot of the wave packet.

We might equally well have chosen a specific position x_0 and measured the time-dependent amplitude $\psi(x_0,t)$ to obtain a function of time only: $\psi(t) := \psi(x_0,t)$. Just as $\psi(x)$ can be written as a superposition of waves, as in Eq. (6.3), so can $\psi(t)$:

$$\psi(t) \ =\ \int A(\omega)e^{i\omega t}\,d\omega\ .$$

The name of the dummy variable of integration has been changed from k to ω in honour of its physical interpretation as an angular frequency, but the mathematics is identical and everything we said about functions of x applies equally well to functions of t. In particular, the bandwidth theorem takes the form

$$\boxed{\Delta t\,\Delta \omega \ \geq\ \frac{1}{2}} \tag{6.5}$$

This version of the bandwidth theorem provides a fundamental limit on the information carrying capacity of any communications system: in order to transmit B bits per second, implying a wavepacket duration $\Delta t \approx 1/B$, you need a bandwidth of at least $\Delta\omega = 1/(2\Delta t) = B/2$. This explains why governments can raise billions by auctioning off regions of the electromagnetic spectrum to communications companies.

6.3 Heisenberg's Uncertainty Principle

In quantum mechanics, de Broglie's equation says that the momentum p is equal to $\hbar k$. Multiplying the wave vector/position version of the bandwidth theorem, Eq. (6.4), by $\hbar$ therefore gives the position-momentum version of Heisenberg's uncertainty principle:

$$\boxed{\Delta x\,\Delta p \;\geq\; \frac{\hbar}{2}} \tag{6.6}$$

To use the uncertainty principle properly, it is important to remember that Δx is the rms width of $\psi(x)$ and $\Delta k = \Delta p/\hbar$ is the rms width of $A(k)$.

Similarly, since Planck's equation says that $E = \hbar\omega$, multiplying the frequency/time version of the bandwidth theorem, Eq. (6.5), by $\hbar$ gives the time-energy version of Heisenberg's uncertainty principle:

$$\boxed{\Delta t\,\Delta E \;\geq\; \frac{\hbar}{2}} \tag{6.7}$$

6.3.1 Practical meaning

The uncertainty principle is easy to misinterpret and many explanations in books are confusing and/or confused. One way to cut through the gobbledegook is to focus on an experiment that could be used to test it.

Start with a large number (say $2N$) of identical quantum mechanical systems, each containing a single quantum mechanical particle. You might, for example, start with $2N$ hydrogen atoms, each containing a single electron in its ground state.

In N of the systems, measure the particle's position: $\{x_i;\ i = 1, N\}$.
In the *other* N, measure the particle's momentum: $\{p_i;\ i = 1, N\}$.

From the measurements of position, evaluate

$$\langle x\rangle \;=\; \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{and} \qquad (\Delta x)^2 \;=\; \langle (x-\langle x\rangle)^2\rangle \;=\; \frac{1}{N}\sum_{i=1}^{N}(x_i-\langle x\rangle)^2.$$

From the measurements of momentum, evaluate

$$\langle p\rangle \;=\; \frac{1}{N}\sum_{j=1}^{N} p_j \qquad \text{and} \qquad (\Delta p)^2 \;=\; \langle (p-\langle p\rangle)^2\rangle \;=\; \frac{1}{N}\sum_{j=1}^{N}(p_j-\langle p\rangle)^2.$$

Then, for large enough N, one always finds that

$$\Delta x\,\Delta p \;\geq\; \frac{\hbar}{2}\,.$$

- The more precisely you can predict the outcome of a measurement of x (p), the less precisely you can predict the outcome of a measurement of p (x). This is very unlike classical physics, where you can in principle know enough about the system to predict the precise values of both x and p.

- Immediately after a measurement of x (p), the value of Δx (Δp) must be very small and hence Δp (Δx) must be very large. This means that you cannot improve your knowledge of (and thus ability to predict the outcome of a measurement of) one of the two without simultaneously decreasing the precision of your knowledge of the other.

6.3.2 The Heisenberg microscope

The Heisenberg microscope was a thought experiment designed to beat the uncertainty principle. Since the uncertainty principle (as described in these notes, at any rate) is a provable statement about the mathematics of Fourier analysis this is a hopeless task, but that need not stop us trying.

The idea is to use the apparatus shown in Fig. 6.4 to image an electron with a known incident momentum p_x^{in}. If we can resolve the position accurately without simultaneously increasing the momentum uncertainty too much, we may end up able to predict the outcomes of future measurements of x and p more precisely than the uncertainty principle allows.

The theory of optics says that the resolution of a lens is limited by diffraction to

$$\Delta x = \frac{\lambda}{\sin\phi}\,.$$

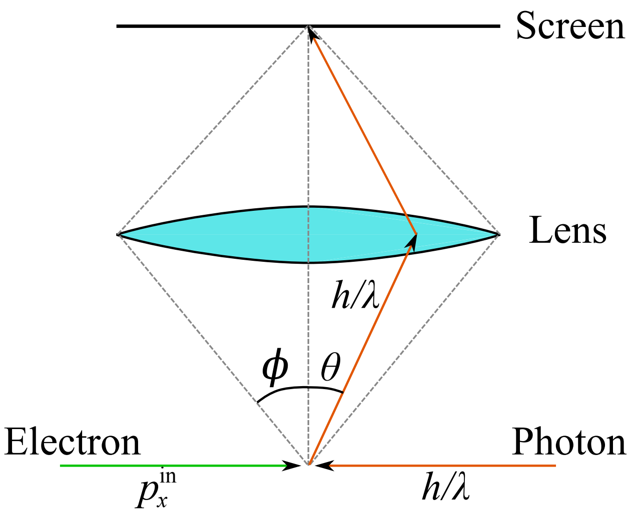


Figure 6.4: The Heisenberg microscope

The scattering of a single photon changes the x component of the electron momentum from p_x^{in} to

$$p_x \;=\; p_x^{\text{in}} - \frac{h}{\lambda} - \frac{h}{\lambda}\sin\theta \;.$$

We know only that $-\phi < \theta < \phi$ and hence that

$$p_x^{\text{in}} - \frac{h}{\lambda} - \frac{h}{\lambda}\sin\phi \;<\; p_x \;<\; p_x^{\text{in}} - \frac{h}{\lambda} + \frac{h}{\lambda}\sin\phi \;.$$

This gives $\Delta p_x \approx \frac{h}{\lambda}\sin\phi$ and

$$\Delta x\,\Delta p_x \;\approx\; \left(\frac{\lambda}{\sin\phi}\right)\left(\frac{h}{\lambda}\sin\phi\right) \;\approx\; h \;.$$

The Heisenberg uncertainty principle survives the test!

6.4 Consequences of the Uncertainty Principle

6.4.1 Diffraction

Why do particles diffract? One way to understand is via the uncertainty principle. On emerging from the slit, the uncertainty Δy in the vertical position of the

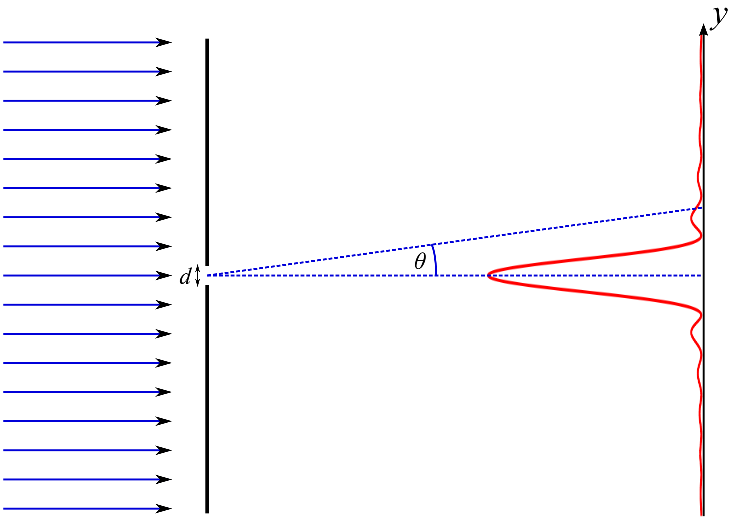
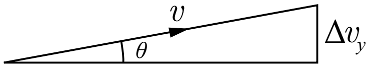


Figure 6.5: Single-slit diffraction

particle is of order $d/2$. Combining this with the Heisenberg uncertainty principle gives

$$\Delta p_y \;\geq\; \frac{\hbar}{2\Delta y} \;\approx\; \frac{\hbar}{d}\,.$$

Assuming that the deviation angle θ is small, so that $\sin\theta = \Delta v_y/v \approx \theta$,



this gives

$$\theta \;\approx\; \frac{\Delta v_y}{v} \;=\; \frac{\Delta p_y}{p} \;\gtrapprox\; \frac{\hbar}{d\hbar k} \;=\; \frac{1}{kd}\,.$$

From problem sheet 1, Q8, the first zero of the diffraction pattern occurs where

$$\frac{kd\sin\theta}{2} = \pi \qquad \Rightarrow \qquad \theta \;\approx\; \frac{2\pi}{kd} \qquad (\text{since } \theta \text{ is small}).$$

The estimate obtained from the uncertainty principle is smaller than this — as it should be.

6.4.2 Zero-point energy

A particle confined in a "box" of size ℓ has

$$\Delta x \;\approx\; \frac{\ell}{2}\,.$$

(Note that we are only after an order of magnitude estimate here; it would have been equally acceptable to use ℓ or $\ell/3$ or ...) Hence

$$\Delta p \;\geq\; \frac{\hbar}{2\Delta x} \;\approx\; \frac{\hbar}{\ell}\,.$$

Since $\langle p\rangle = 0$ (the particle is confined in the box, so its average momentum must be zero), $(\Delta p)^2 = \langle (p - \langle p\rangle)^2\rangle = \langle p^2\rangle$. Hence

$$\langle KE\rangle \;=\; \frac{\langle p^2\rangle}{2m} \;=\; \frac{(\Delta p)^2}{2m} \;\gtrapprox\; \frac{\hbar^2}{2m\ell^2}\,.$$

Decreasing ℓ increases this "zero-point" kinetic energy.

The existence of zero-point energy explains:

- Why the electrons in atoms do not spiral into the nucleus. When ℓ is as small as a nucleus, the zero-point energy cost of confining the electron, which is proportional to $1/\ell^2$, outweighs the Coulomb potential energy gain, which is proportional to $1/\ell$. The kinetic energy cost of confining protons and neutrons is much smaller because they are much heavier than electrons; the forces binding them together (the strong interaction) are also much stronger.

- Why the motion of electrons in atoms and atoms in molecules and solids does not cease as $T \to 0$.

- Why quantum mechanical particles suddenly released from boxes such as atomic traps spread out.

$$\Delta p \;\gtrapprox\; \frac{\hbar}{\ell} \qquad \Rightarrow \qquad \Delta v \;\gtrapprox\; \frac{\hbar}{m\ell} \qquad \text{(typical spreading speed)}.$$

6.4.3 Short-lived quantum states

Because of the time-energy version of the uncertainty principle,

$$\Delta t\,\Delta E \;\geq\; \frac{\hbar}{2}\,,$$

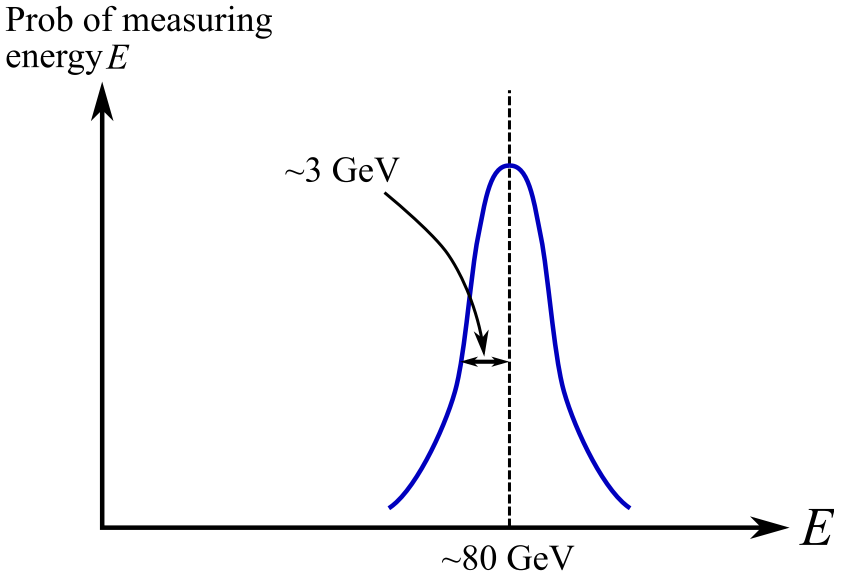


Figure 6.6: The width of the W.

the wave packets associated with short-lived quantum states have uncertain energies.

For example, the W particle (the charged intermediate vector boson responsible, with the Z, for the weak interaction) has a mass of order 80 GeV/c^2 and a lifetime of order 10^{-25}s. How precisely can its energy be specified?

$$\Delta E \gtrapprox \frac{\hbar}{2\Delta t} = \frac{1.05\times 10^{-34}}{2\times 10^{-25}} \approx 5.25\times 10^{-10}\,\mathrm{J} \approx 3.3\,\mathrm{GeV}.$$

This is known as the width of the W.

Another consequence of the time-energy uncertainty principle is that things with very short lifetime do not have to conserve energy exactly. “Virtual” particles can appear out of the vacuum briefly before disappearing again (and they do!).

6.5 Group Velocity

You already know that the envelope of a “musical” wave packet (that is, a wave packet containing a narrow range of k vectors centred on the carrier wave vector k_c) travels at the group velocity, but may not have been told why. In case you are interested, a proof based on the idea of Fourier superposition is included below. *This section is not examinable.*

6.5.1 Two-Wave Example

To introduce the main ideas, let us return to the two-wave example discussed earlier. Since we are interested in how the wave packet moves, we have to consider its time-dependence and position-dependence together. Equation (6.2) is therefore replaced by

$$\psi(x,t) = \psi_1(x,t) + \psi_2(x,t) = \cos(k_1 x - \omega_1 t) + \cos(k_2 x - \omega_2 t) \ ,$$

which describes a superposition of two travelling waves with slightly different wavelengths. The angular frequencies ω_1 and ω_2 are related to k_1 and k_2 via the dispersion relation $\omega = \omega(k)$.

As t increases, the two cosine functions move to the right at the phase velocity. For light waves with dispersion relation $\omega = ck$, the phase velocities

$$v_1 \ = \ \frac{\omega_1}{k_1} \ = \ \frac{ck_1}{k_1} \ = \ c$$

and

$$v_2 \ = \ \frac{\omega_2}{k_2} \ = \ \frac{ck_2}{k_2} \ = \ c$$

are both equal to c. Since the interference pattern of beats is just the sum of ψ_1 and ψ_2, both of which are moving at speed c, it also moves at speed c. Apart from this constant motion, the shape of the interference pattern never changes.

For other kinds of waves the dispersion relation is more complicated and v_1 and v_2 may differ. This makes it much harder to figure out how the interference pattern of beats moves and changes with time. In fact, the beat pattern moves at the group velocity $d\omega/dk$ rather than the phase velocity ω/k. For light waves, $d\omega/dk = \omega/k = c$ and so the group and phase velocities are the same. For de Broglie particle-waves, which have the dispersion relation $\omega = \hbar k^2/2m$, the group velocity

$$\frac{d\omega}{dk} = \frac{\hbar k}{m}$$

is twice the phase velocity

$$\frac{\omega}{k} = \frac{\hbar k}{2m} \ .$$

This means that the beat pattern, created by the interference of the two cosine waves, moves twice as fast as the waves themselves.

To see why the interference pattern moves at the group velocity, consider the regions where the two components of

$$\psi(x,t) = \cos(k_1 x - \omega_1 t) + \cos(k_2 x - \omega t)$$

interfere contructively. This happens where $k_1 x - \omega_1 t$ and $k_2 x - \omega_2 t$ differ by a multiple of 2π:

$$k_1 x - \omega_1 t = k_2 x - \omega_2 t + 2\pi n \qquad (n \text{ any integer})\ .$$

For the $n = 0$ peak, the condition for constructive interference reduces to:

$$(k_2 - k_1)x = (\omega_2 - \omega_1)t\ .$$

When $t = 0$, the solution of this equation is $x = 0$ (in other words, the $n = 0$ peak is the broad peak in the middle of the lower panel in Fig. 6.3). When $t > 0$, the position of the $n = 0$ peak is given by:

$$x = \frac{\omega_2 - \omega_1}{k_2 - k_1} t = \frac{\Delta\omega}{\Delta k} t\ .$$

The central peak of the interference pattern therefore moves at speed $\Delta\omega/\Delta k$.

If the wavelengths (and hence wave vectors and angular frequencies) of the two component waves are similar enough, the fraction $\Delta\omega/\Delta k$ is approximately equal to the derivative $d\omega/dk$ (see Fig. 6.7). The interference pattern therefore moves at the group velocity:

$$v_g(k) = \frac{d\omega}{dk}\ .$$

Since Δk is assumed to be very small, it makes little difference whether the group velocity is evaluated at k_1 or k_2: $v_g(k_1) \approx v_g(k_2) \approx v_g((k_1 + k_2)/2)$. For aesthetic reasons, I prefer to use the average wave vector $(k_1 + k_2)/2$.

If the wavelengths of the two component waves differ by too much, the approximation $\Delta\omega/\Delta k \approx d\omega/dk$ may be poor. In this case, the velocity $\Delta\omega/\Delta k$ of the interference envelope will not be the same as the group velocity $d\omega/dk$.

6.5.2 General Case

Now consider a general wave packet constructed by superposing many cosine waves. If the wave packet is a "musical" one — in other words, if all of the

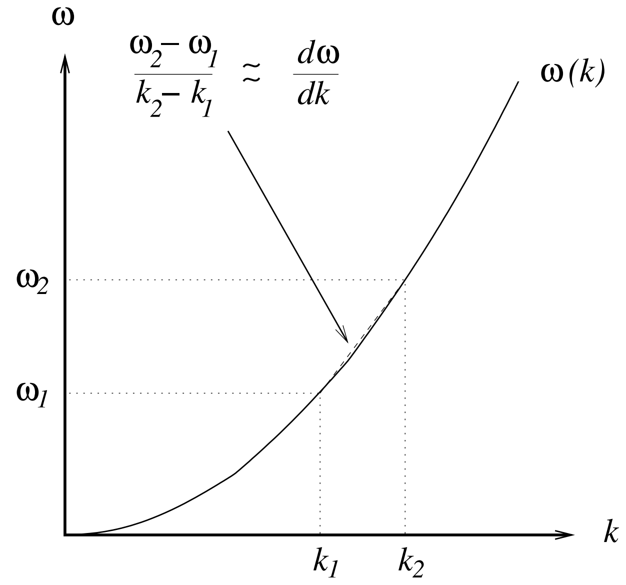


Figure 6.7: The slope $(\omega_2 - \omega_1)/(k_2 - k_1)$ of the (almost invisible) dashed line is approximately the same as the slope $d\omega/dk$ of the function $\omega(k)$.

waves contributing to the superposition have similar wavelengths — it is possible to show that the interference envelope moves at the group velocity.

The proof uses the complex representation

$$\psi(x,t) = \sum_{n=1}^{N} A_n e^{i(k_n x - \omega_n t)} \ ,$$

of a wave packet constructed by superposing many travelling waves. As usual, the "Re" symbol has been omitted.

Since the wave packet is musical, the wave vectors k_n are all very close to the carrier wave vector k_c:

$$k_n = k_c + \Delta k_n \ ,$$

where Δk_n is small. The angular frequency $\omega_n = \omega(k_n)$ may therefore be approximated using the first two terms of a Taylor series:

$$\omega_n \ = \ \omega(k_c + \Delta k_n) \ \approx \ \omega(k_c) + \left.\frac{d\omega}{dk}\right|_{k=k_c} \Delta k_n \ = \ \omega_c + v_g \Delta k_n \ ,$$

where $v_g = d\omega/dk|_{k=k_c}$. The expression for $\psi(x,t)$ then becomes:

$$\begin{aligned}\psi(x,t) &\approx \sum_{n=1}^{N} A_n e^{i[(k_c+\Delta k_n)x-(\omega_c+v_g\Delta k_n)t]} \\ &= \sum_{n=1}^{N} A_n e^{i(k_c x-\omega_c t)+i\Delta k_n(x-v_g t)} \\ &= e^{i(k_c x-\omega_c t)} \sum_{n=1}^{N} A_n e^{i\Delta k_n(x-v_g t)} \ .\end{aligned}$$

The exponential prefactor is the carrier wave with wavelength $\lambda_c = 2\pi/k_c$, while the summation gives the shape of the envelope. The important point is that the envelope is a function of $x - v_g t$ only:

$$\psi(x,t) \approx e^{i(k_c x-\omega_c t)} f(x - v_g t) \ .$$

This means that the envelope has the same shape [the shape of $f(x)$] at all times. As t increases, this frozen shape simply moves to the right at speed v_g.

The only approximation in the above derivation was the replacement of ω_n by the first two terms of a Taylor series. This approximation is exact if the dispersion relation is linear ($\omega = ck$, as for light) and good whenever the dispersion relation is close to linear over the range of wave vectors contributing to the wave packet. If the dispersion relation is not quite linear, however, the neglected higher-order terms cause the wave packet to smear out as it moves along. The wider the spread of wavelengths in the wave packet, the more rapidly this smearing (called dispersion) occurs.

Chapter 7

The Schrödinger Equation

7.1 The Time-Dependent Schrödinger Equation

> Once at the end of a colloquium I head Debye saying something like: “Schödinger, you are not working right now on very important problems anyway. Why don’t you tell us some time about that thesis of de Broglie, which seems to have attracted some attention?” So, in one of the next colloquia, Schrödinger gave a beautifully clear account of how de Broglie associated a wave with a particle and how he could obtain the quantization rules of Neils Bohr and Sommerfeld by demanding that an integer number of waves should be fitted along a stationary orbit. When he had finished, Debye casually remarked that he thought this way of talking was rather childish. As a student of Sommerfeld he had learned that, to deal properly with waves, one had to have a wave equation. It sounded quite trivial and did not seem to make a great impression, but Schödinger evidently thought a bit more about the idea afterwards.
>
> F. Bloch, “Reminiscences of Heisenberg and the early days of quantum mechanics”, *Physics Today* **29** (12), pp. 23–27 (1976).

Some books try to derive Schrödinger’s wave equation by starting from Newton’s second law and applying mysterious quantisation rules. This is philosophical nonsense. Schrödinger’s equation provides a more fundamental and widely applicable description of the world than Newton’s second law (which does not work at the atomic scale) and cannot be “derived” from classical mechanics. If anything, Newton’s second law ought to be regarded as an approximation to the underlying quantum theory — an approximation that works very well when applied to large objects but fails miserably on the atomic scale.

Although it is not possible to derive the Schrödinger equation (or any other fundamental law of nature), the experimental evidence discussed earlier in this course suggests that it ought to have the following properties:

(a) Linearity

> If $\psi_1(x,t)$ and $\psi_2(x,t)$ are solutions, then $b_1\psi_1(x,t) + b_2\psi_2(x,t)$ must also be a solution for any choice of the constants b_1 and b_2.
>
> This reflects our understanding that quantum mechanical waves superpose like "normal" waves; their amplitudes add when they overlap, but they pass through each other unaltered, evolving as if the other overlapping waves were not there. The linearity condition also implies that the normalisation of the wave function is not a property of the Schrödinger equation itself: if $\psi(x,t)$ is a solution then so is $b\psi(x,t)$ for any constant b. The normalistion has to be imposed "by hand".

(b) Dispersion relation

> For a free particle (one that experiences no forces), the wave function $e^{i(kx-\omega t)}$ ought to be a solution of the Schrödinger equation if and only if
>
> $$\frac{\hbar^2 k^2}{2m} = \hbar\omega \qquad \left(\frac{p^2}{2m} = E\right).$$

Solutions of the ordinary wave equation,

$$c^2 \frac{\partial^2 \psi}{\partial x^2} = \frac{\partial^2 \psi}{\partial t^2} ,$$

satisfy (a) but not (b).

- To verify that (a) is satisfied, suppose that ψ_1 and ψ_2 are two solutions of the ordinary wave equation:

  $$c^2 \frac{\partial^2 \psi_1}{\partial x^2} = \frac{\partial^2 \psi_1}{\partial t^2} , \qquad c^2 \frac{\partial^2 \psi_2}{\partial x^2} = \frac{\partial^2 \psi_2}{\partial t^2} .$$

  Adding b_1 times the first equation to b_2 times the second gives

  $$b_1 c^2 \frac{\partial^2 \psi_1}{\partial x^2} + b_2 c^2 \frac{\partial^2 \psi_2}{\partial x^2} = b_1 \frac{\partial^2 \psi_1}{\partial t^2} + b_2 \frac{\partial^2 \psi_2}{\partial t^2} ,$$

and hence

$$c^2\frac{\partial^2(b_1\psi_1+b_2\psi_2)}{\partial x^2} \;=\; \frac{\partial^2(b_1\psi_1+b_2\psi_2)}{\partial t^2}\;.$$

This shows that $b_1\psi_1 + b_2\psi_2$ also satisfies the ordinary wave equation.

- To verify that (b) is *not* satisfied, note that:

  $$c^2\frac{\partial^2 e^{i(kx-\omega t)}}{\partial x^2} \;=\; -c^2k^2e^{i(kx-\omega t)} \quad\text{and}\quad \frac{\partial^2 e^{i(kx-\omega t)}}{\partial t^2} \;=\; -\omega^2 e^{i(kx-\omega t)}\;.$$

  Hence, $e^{i(kx-\omega t)}$ satisfies the ordinary wave equation if and only if

  $$\omega^2 = c^2k^2 \qquad\Rightarrow\qquad \omega = \pm ck\;.$$

  This is the correct dispersion relation for light waves or waves on a violin string, but no good for wavicles (which have $\omega = \hbar k^2/2m$).

The ordinary wave equation may not produce the correct dispersion relation for wavicles, but its analysis suggests how to write down an equation that does. If we assume that the Schrödinger equation for a free particle has solutions of the form $\psi(x,t) = e^{i(kx-\omega t)}$, every application of $\partial/\partial x$ pulls down a factor of ik and every application of $\partial/\partial t$ pulls down a factor of $-i\omega$. To reproduce the wavicle dispersion relation, $\hbar\omega = \hbar^2k^2/2m$, requires an equation with one t derivative and two x derivatives. To get rid of the $-i$ factor introduced by the time derivative, the equation also has to include an i.

This reasoning leads directly to the *time-dependent Schrödinger equation for a free particle*:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} \;=\; i\hbar\frac{\partial\psi}{\partial t}\;. \tag{7.1}$$

Like the ordinary wave equation, the Schrödinger equation is linear. Unlike the ordinary wave equation, it leads to the correct dispersion relation for free wavicles. To check this, note that

$$-\frac{\hbar^2}{2m}\frac{\partial^2 e^{i(kx-\omega t)}}{\partial x^2} \;=\; \frac{\hbar^2k^2}{2m}e^{i(kx-\omega t)} \quad\text{and}\quad i\hbar\frac{\partial e^{i(kx-\omega t)}}{\partial t} = \hbar\omega e^{i(kx-\omega t)}\;.$$

Hence $e^{i(kx-\omega t)}$ satisfies Eq. (7.1) if and only if $\hbar\omega = \hbar^2k^2/2m$.

Eq. (7.1) works for free particles (on which no forces act), but not for particles moving through a potential $V(x)$ and experiencing forces $F(x) = -dV(x)/dx$.

The total energy of such a particle is $p^2/2m + V(x)$, suggesting the dispersion relation

$$\frac{\hbar^2 k^2}{2m} + V(x) = \hbar\omega$$

and the corresponding wave equation

$$\boxed{-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial x^2} + V(x)\psi(x,t) \;=\; i\hbar\frac{\partial\psi(x,t)}{\partial t}} \qquad (7.2)$$

(A ψ factor has to be included in the potential energy term to keep the equation linear.) Equation (7.2) is the general form of the *time-dependent Schrödinger equation* for one particle in one dimension.

Aside: Potentials in QM and E&M

In electricity and magnetism courses, the "potential at x" is the potential energy of a unit test charge placed at x. In quantum mechanics, a slightly different definition is used: the "potential at x" is not the potential energy of a test charge but the potential energy of the physical particle being studied at x. For example, in an E&M course, the potential a distance r from a nucleus of charge Ze would be $Ze/4\pi\epsilon_0 r$; in a QM course, the potential of an electron (charge $-e$) a distance r from the nucleus would be $-Ze^2/4\pi\epsilon_0 r$.

Notes

- We note again that we have not *derived* Eq. (7.1) or Eq. (7.2); we merely *proposed* them as the simplest differential equations we could think of consistent with the experimental results.

- Because of the i factor, the quantum mechanical wave function $\psi(x,t)$ is *necessarily* complex. Even if $\psi(x,t)$ starts off real, the time evolution described by Eq. (7.2) soon makes it complex.

- It can be shown that Eq. (7.2) conserves (but of course does not fix) the normalisation of $\psi(x,t)$:

  $$\frac{\partial}{\partial t}\int_{-\infty}^{\infty}\psi^*(x,t)\psi(x,t)\,dx \;=\; 0 \quad \text{if} \quad \psi(x,t)\to 0 \;\text{ as }\; |x|\to\infty.$$

  (You can prove this by moving the time derivatives inside the integral and using Eq. (7.2) to substitute for $\partial\psi^*/\partial t$ and $\partial\psi/\partial t$. The potential energy contributions from the $\partial\psi^*/\partial t$ and $\partial\psi/\partial t$ terms cancel, and integration by parts can be used to show that the kinetic energy terms also vanish. Try it.)

- Equation (7.2) is often written as

$$\hat{H}\psi \;=\; i\hbar\frac{\partial\psi}{\partial t}\,, \tag{7.3}$$

where

$$\hat{H} \;:=\; -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) \tag{7.4}$$

is known as the *Hamiltonian operator* or simply the *Hamiltonian*.

- Since, for small enough Δt,

$$\frac{\partial\psi(x,t)}{\partial t} \;\approx\; \frac{\psi(x,t+\Delta t)-\psi(x,t)}{\Delta t}\,,$$

Eq. (7.2) is equivalent to

$$i\hbar\psi(x,t+\Delta t) \;\approx\; i\hbar\psi(x,t) + \Delta t\left(-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial x^2} + V(x)\psi(x,t)\right)\,.$$

Given the wave function $\psi(x,0)$ at time t=0, this equation allows you to step forward to find $\psi(x,\Delta t)$, then $\psi(x,2\Delta t)$, and so on. The mathematical evolution described by the Schrödinger equation is therefore entirely deterministic: given the wave function at time t=0, you can predict it at all future times. In this sense (ignoring the complications of the poorly understood measurement process), Schrödinger's universe is just as "clockwork" as Newton's. The difference is that Schrödinger's clockwork evolution applies only to *probability densities* for the outcomes of measurements and does not in general allow you to make certain predictions about the future.

7.2 Energy Levels

The time-dependent Schrödinger equation tells you $\psi(x,t)$ if you know $\psi(x,0)$, but that is all it does; it is an *equation of motion* and does not say anything (directly) about quantisation or energy levels. For example, as illustrated in Fig. 7.1, the wave function $\psi(x,t)$ of a particle in a box could be more or less anything normalised and consistent with the boundary conditions; it does not have to look like one of the standing waves on a violin string.

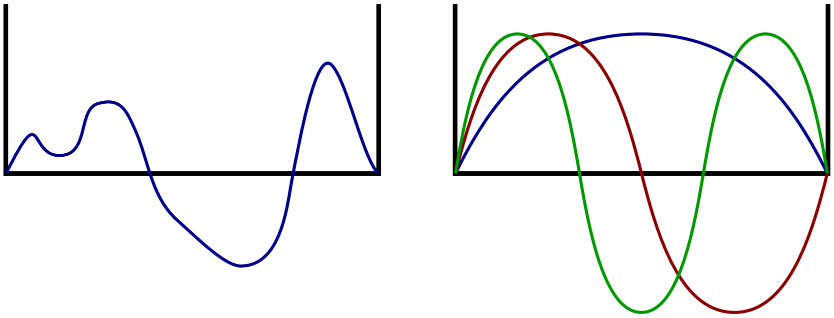


Figure 7.1: The left-hand figure shows a snapshot of the real part of a typical wave function for a particle in a box. It is normalised, satisfies the boundary conditions ($\psi = 0$ at the box sides), and evolves according to the time-dependent Schrödinger equation, but that is more or less all you can say about it in general; its detailed form depends on the history of the system. The right-hand figure shows the standing waves that correspond to the first three quantised energy levels and satisfy the time-independent Schrödinger equation derived in Sec. 7.2.

Similar considerations apply to the wave function of a free particle. By construction (see conditions (a) and (b) in Sec. 7.1), the time-dependent Schrödinger equation for a free particle, Eq. (7.1), is satisfied by any wave function of the form

$$\psi(x,t) = \sum_n A_n e^{i(k_n x - \omega_n t)} \ ,$$

where $\hbar\omega_n = \hbar^2 k_n^2/2m$. Since this wave function contains various angular frequencies ω_n, it cannot have a precise energy. If you measure the energy, you may obtain any of the values $E_n = \hbar\omega_n$ for which $|A_n|^2$ is non-zero.

A wave function with a precise energy, often known as an *energy eigenfunction*, is any solution of the time-dependent Schrödinger equation involving only a single angular frequency. Such wave functions are directly analogous to the standing waves on a violin string, which also have precise frequencies (and hence pitches). For a free particle, the quantum mechanical energy eigenfunctions are pure travelling waves such as

$$\psi(x,t) = e^{i(kx-\omega t)} \qquad \text{with} \qquad \hbar\omega = \hbar^2 k^2/2m \, .$$

If you measure the energy of a free particle described by a wave function like this, the only possible answer is $\hbar\omega$. The energy uncertainty ΔE is therefore zero. Fortunately, since $e^{i(kx-\omega t)}$ does not die off as $t \to \pm\infty$, the time uncertainty

Δt is infinite. This means that the uncertainty product $\Delta t\,\Delta E$ is undefined. The uncertainty principle (which only really applies to finite wave packets) cannot be used, but at least it is not obviously disobeyed.

The energy eigenfunctions of confined particles such as the electrons in atoms or the particle in a box studied in Sec. 5.4 are more interesting. Substituting a single-frequency trial solution of the form

$$\psi(x,t) = \phi(x)e^{-i\omega t}$$

into the time-dependent Schrödinger equation, Eq. (7.2), gives

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\left(\phi e^{-i\omega t}\right) + V\left(\phi e^{-i\omega t}\right) \;=\; i\hbar\frac{\partial}{\partial t}\left(\phi e^{-i\omega t}\right)\ .$$

Since ϕ depends only on x and $e^{-i\omega t}$ depends only on t, this simplifies to

$$e^{-i\omega t}\left(-\frac{\hbar^2}{2m}\frac{\partial^2\phi}{\partial x^2}\right) + e^{-i\omega t}V\phi \;=\; e^{-i\omega t}\hbar\omega\phi\ .$$

Cancelling the common $e^{-i\omega t}$ factor (which can never be zero) gives the *time-independent Schrödinger equation*,

$$-\frac{\hbar^2}{2m}\frac{\partial^2\phi(x)}{\partial x^2} + V(x)\phi(x) \;=\; E\phi(x) \tag{7.5}$$

where $E = \hbar\omega$. This may also be written in terms of the Hamiltonian operator defined in Eq. (7.4):

$$\hat{H}\phi(x) \;=\; E\phi(x)\ . \tag{7.6}$$

There is an obvious similarity between Eq. (7.6) and a matrix eigenvalue problem such as

$$\mathbf{M}\mathbf{v} = \lambda\mathbf{v}\ ,$$

where M is a square matrix, λ an eigenvalue, and **v** an eigenvector. This explains why E is called an energy eigenvalue and ϕ an energy eigenfunction.

For confined or bound particles, such as electrons in atoms, the solutions of Eq. (7.5) are discrete and can be labelled using integers:

$$\left.\begin{array}{l}\phi_n(x)\\ E_n\end{array}\right\}\qquad n = 1, 2, 3, \ldots$$

- The energy eigenfunctions $\phi_n(x)$ are direct mathematical and physical analogues of the standing waves on a violin string; the angular frequencies $\omega_n = E_n/\hbar$ are analogous to violin harmonics.
- The eigenvalues E_n are the quantised energy levels observed in experiments. Except for small relativistic corrections, the eigenvalues obtained by solving the time-independent Schrödinger equation give the energy levels of electrons in atoms, molecules and solids *exactly*.
- The time-independent Schrödinger equation predicts the shell structure of atoms and hence explains the form of the periodic table of elements. In fact, in principle at least, the Schrödinger equation explains *all* of chemistry and *all* of the physics of everyday matter.

The mathematical theory of the time-independent Schrödinger equation is straightforward and well understood, but actually calculating the eigenvalues and eigenfunctions of complicated systems of many particles remains difficult. A few simple examples will be covered next year, but the process normally requires a computer if it can be done at all. For this course, you will only be asked to verify that a given trial solution is an eigenfunction and to find the corresponding eigenvalue.

7.2.1 The square well

The simplest interesting example is the particle in a box introduced in Sec. 5.4. The box and the first three energy eigenfunctions are shown in Fig. 7.2. The time-independent Schrödinger equation for the n^{th} eigenfunction $\phi_n(x)$ is

$$-\frac{\hbar^2}{2m}\frac{d^2\phi_n(x)}{dx^2} + V(x)\phi_n(x) = E_n\phi_n(x) \ ,$$

where

$$V(x) \ = \ \begin{cases} 0 & 0 < x < a, \\ \infty & \text{otherwise.} \end{cases}$$

The process of verifying that $\phi_n(x) = \sqrt{2/a}\ \sin(n\pi x/a)$ is a normalised energy eigenfunction and finding the corresponding energy eigenvalue has three steps.

1. Check that the trial eigenfunction satisfies the boundary conditions

   $$\phi_n(0) = \phi_n(a) = 0 \ .$$

   Since $\sin(n\pi) = 0$ for any integer n, this is obviously the case.

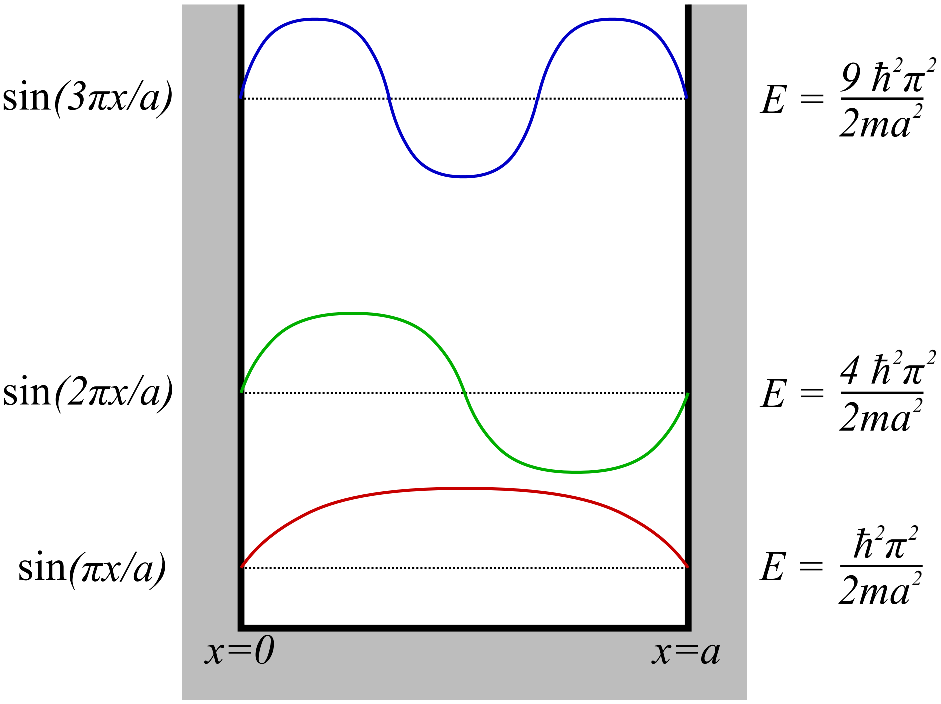


Figure 7.2: A square well of width a, showing the first three energy eigenfunctions (not normlalised) and the corresponding energy eigenvalues.

2. Check that the Schrödinger equation is satisfied in the region $0 < x < a$:

$$\begin{aligned}
&\left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x)\right)\phi_n(x) \\
&= -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\left(\sqrt{\frac{2}{a}}\sin\left(\frac{n\pi x}{a}\right)\right) \qquad \text{(since } V(x)\text{=0 if } 0 < x < a\text{)} \\
&= \frac{n^2\hbar^2\pi^2}{2ma^2}\sqrt{\frac{2}{a}}\sin\left(\frac{n\pi x}{a}\right) \\
&= \frac{n^2\hbar^2\pi^2}{2ma^2}\phi_n(x)\ .
\end{aligned}$$

Hence $\phi_n(x)$ satisfies the time-independent Schrödinger equation with the eigenvalue

$$E_n = \frac{n^2\hbar^2\pi^2}{2ma^2}\ .$$

The ground state of a *classical* particle in a box has zero kinetic energy, zero potential energy, and zero total energy. In the quantum mechanical version, the ground state $\phi_1(x)$ has energy $E_1 = \hbar^2\pi^2/2ma^2$. The zero-point energy of the particle (the extra quantum mechanical energy that would be missing in classical physics) is therefore $\hbar^2\pi^2/2ma^2$. If the box is made smaller (a decreases), the zero-point energy increases.

3. Check the normalisation:

$$\int_{-\infty}^{\infty} |\phi_n(x)|^2 dx \;=\; \int_0^a \frac{2}{a}\sin^2\left(\frac{n\pi x}{a}\right) dx \;=\; 1 \qquad \text{(exercise).}$$

7.2.2 The quantum mechanical simple harmonic oscillator

The potential energy of a classical simple harmonic oscillator is

$$V(x) = \frac{1}{2}sx^2 \ ,$$

where x is the displacement of the particle from its equilibrium position and s is the spring constant. The corresponding expression for the force,

$$F(x) = -\frac{dV}{dx} = -sx \ ,$$

is known as Hooke's law. The motion of the classical particle is determined by Newton's laws.

A quantum mechanical simple harmonic oscillator is any system in which a quantum mechanical particle feels a potential of the form $\frac{1}{2}sx^2$. For small displacements, the potential energy of an atom vibrating about its equilibrium position in a molecule or solid is well approximated by a quadratic of this type. Because atoms are small enough for quantum effects to be important in some circumstances, their vibrations often need to be treated quantum mechanically.

The energy levels E_n and eigenfunctions $\phi_n(x)$ of the oscillating particle are found by solving the time-independent Schrödinger equation:

$$-\frac{\hbar^2}{2m}\frac{d^2\phi_n(x)}{dx^2} + \frac{1}{2}sx^2\phi_n(x) = E_n\phi_n(x) \ .$$

You will learn how to solve this equation using series expansions next year. For now, all you need to know is the result,

$$E = \left(n + \tfrac{1}{2}\right)\hbar\omega_{\text{cl}} \ , \quad n = 0, 1, 2, \ldots \ ,$$

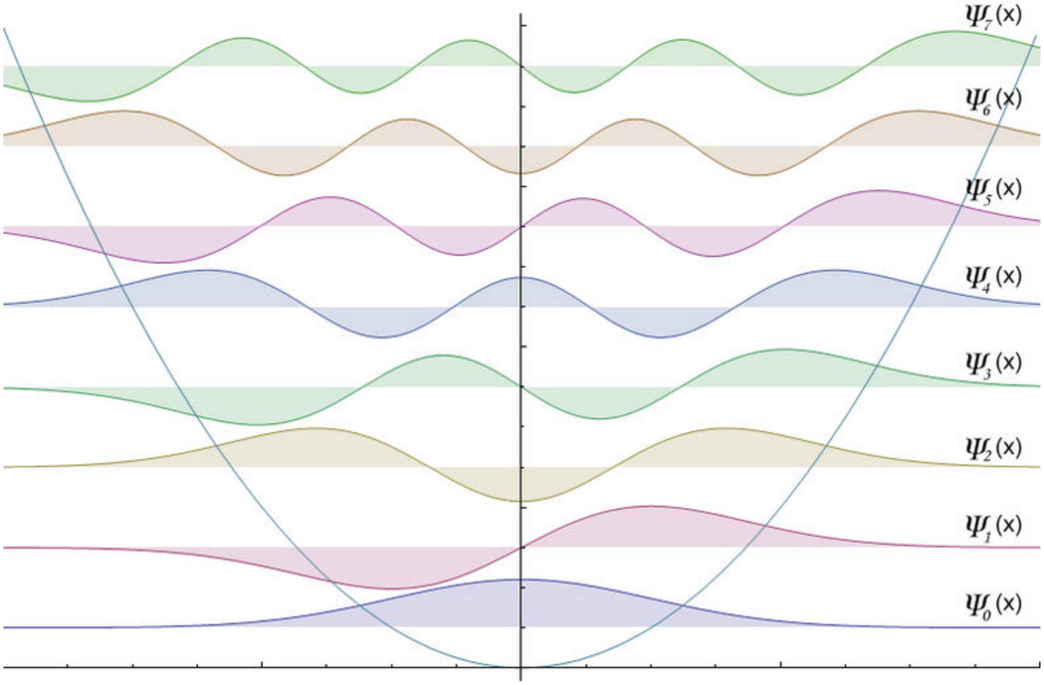


Figure 7.3: The first few energy eigenfunctions of a quantum mechanical simple harmonic oscillator.

where $\omega_{\rm cl} = \sqrt{s/m}$ is the angular frequency of a classical oscillator with spring constant s and mass m. The lowest few energy levels are

$$\begin{array}{lcll} E_0 & = & \frac{1}{2}\hbar\omega_{\rm cl} & \text{(ground state)} \\ E_1 & = & \frac{3}{2}\hbar\omega_{\rm cl} & \text{(first excited state)} \\ E_2 & = & \frac{5}{2}\hbar\omega_{\rm cl} & \text{(second excited state)} \\ \dots & . & \dots & \end{array}$$

The eigenfunctions shown in Fig. 7.3 look similar to those of a particle in a box, although the formulae are considerably more complicated. Since the potential rises smoothly as x increases, there are no “hard wall” boundary conditions forcing the eigenfunctions to zero at specific points. Rather, the eigenfunctions tend smoothly to zero as $|x|$ increases, satisfying the boundary conditions $\phi_n(x) \to 0$ as $x \to \pm\infty$.

It is a special feature of the quantum mechanical simple harmonic oscillator that adjacent energy levels are always separated by the same energy difference

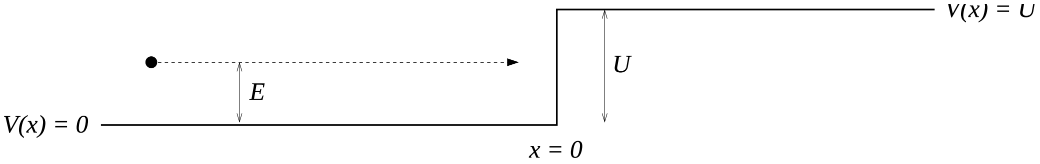
$\hbar\omega_{\mathrm{cl}}$. When molecules switch between adjacent vibrational energy levels, they emit or absorb photons of this energy. Since the frequencies of the modes of oscillation are characteristic of the vibrating molecule, measurements of the energies of the photons emitted and absorbed can be used as a "fingerprint" to identify different molecules. These photons are normally in the infra-red.

The ground state (lowest possible energy state) of a *classical* simple harmonic oscillator has the particle sitting stationary at $x = 0$; this state has zero kinetic energy, zero potential energy, and zero total energy. In a quantum mechanical oscillator, the ground state $\phi_0(x)$ has energy $E_0 = \frac{1}{2}\hbar\omega_{\mathrm{cl}}$. The zero-point energy is therefore $\frac{1}{2}\hbar\omega_{\mathrm{cl}}$.

7.2.3 Tunnelling

Introduction

What happens when a particle of kinetic energy E collides with a potential barrier of height $U > E$?



Since the kinetic energy of the arriving particle is smaller than the potential energy U required to climb the barrier, classical physics says that the particle will always bounce off. In quantum mechanics, by contrast, there is a small chance of finding the particle *inside* the barrier; the probability density dies off exponentially with distance into the barrier, but it is not zero. If the barrier has a finite width, it is even possible for the particle to tunnel through the classically forbidden region and emerge on the other side.

After wading through the mathematics below you may decide that tunnelling is a load of theoretical nonsense. You would be wrong. In the Sun, pairs of positively charged nuclei have to tunnel through the Coulomb potential energy barrier that separates them before fusing to release the energy that powers the Earth. In a scanning tunnelling microscope, measurements of the tiny current that flows as electrons tunnel across the gap between a sharp tip and the surface of a solid are used to build up an atomic-scale relief map of the surface. The astonishing STM image shown in Fig. 7.4 shows what can be done.

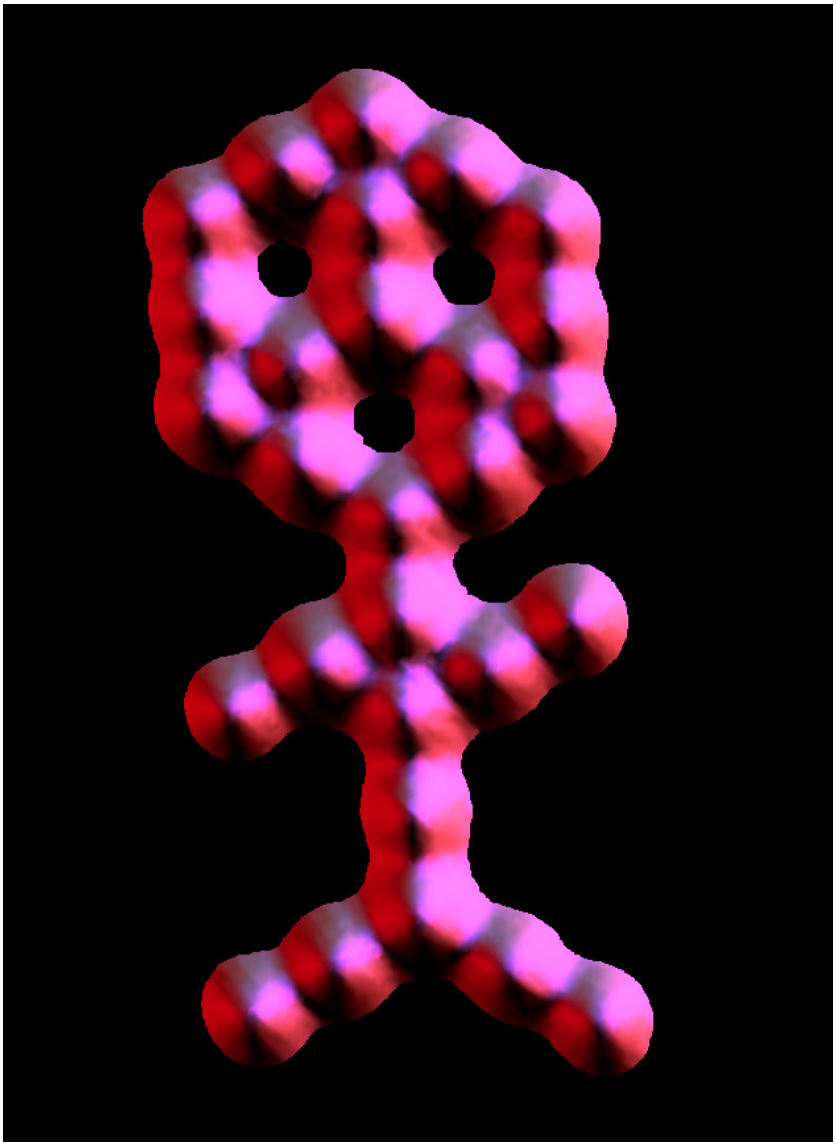


Figure 7.4: A sculpture made by using an STM tip to arrange individual CO molecules on a Pt surface. The STM tip used to move the molecules was also used to image them. [Image originally created by IBM Corporation, http://www.almaden.ibm.com/vis/stm/gallery.html.]

Solutions of the Schrödinger equation

The wave function of a particle with a precise energy $E = \hbar\omega$ has the form $\psi(x,t) = \phi(x)e^{-i\omega t}$. As usual, the spatial part of this wave function satisfies the time-independent Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x)\right]\phi(x) = E\phi(x) \; .$$

Since $V(x) = 0$ if $x < 0$ and $V(x) = U$ if $x > 0$, we can rewrite this equation as follows:

$$\begin{cases} -\dfrac{\hbar^2}{2m}\dfrac{d^2\phi(x)}{dx^2} = E\phi(x) & x < 0 \; , \\[2ex] -\dfrac{\hbar^2}{2m}\dfrac{d^2\phi(x)}{dx^2} = (E-U)\phi(x) & x > 0 \; . \end{cases}$$

These are second-order ordinary differential equations with constant coefficients, so the two independent solutions of each equation are easy to guess.

Solutions for $x < 0$: If we substitute the trial solution $e^{\pm ikx}$ into the left-hand side of the Schrödinger equation for $x < 0$, we obtain

$$-\frac{\hbar^2}{2m}\frac{d^2 e^{\pm ikx}}{dx^2} = \frac{\hbar^2 k^2}{2m}e^{\pm ikx} \; .$$

Thus, the trial solution satisfies the Schrödinger equation if and only if

$$E = \frac{\hbar^2 k^2}{2m} \quad \Rightarrow \quad k = \pm\sqrt{\frac{2mE}{\hbar^2}} \; .$$

The general solution in this region is

$$\phi(x) = Ae^{ikx} + Be^{-ikx} \; ,$$

where A and B are arbitrary (complex) constants. In physical terms, A and B are the complex amplitudes of the incident and reflected waves.

Solutions for $x > 0$: If we substitute the $e^{\pm ikx}$ trial solution into the Schrödinger equation for $x > 0$, we obtain

$$\frac{\hbar^2 k^2}{2m}e^{\pm ikx} = (E-U)e^{\pm ikx} \; ,$$

and hence

$$k = \pm\sqrt{\frac{2m(E-U)}{\hbar^2}} \ .$$

So far, this looks very like the $x < 0$ case. However, because $E - U < 0$, the constant k must now be imaginary:

$$k = \pm i\gamma \qquad \text{with} \qquad \gamma = \sqrt{\frac{2m(U-E)}{\hbar^2}} \ .$$

The general solution in the $x > 0$ region is thus

$$\phi(x) = Ce^{-\gamma x} + De^{\gamma x} \ ,$$

where C and D are arbitrary (complex) constants.

If the barrier is infinitely wide, we can argue on physical grounds that D must be zero (because the probability density $|\phi(x)|^2$ must tend to zero, not infinity, as $x \to \infty$). If the barrier is finite, neither C nor D is exactly zero, but the value of D decreases so rapidly as the barrier widens that the $De^{\gamma x}$ term can often be ignored.

Finding A, B and C for an infinitely wide barrier

The constant A is the amplitude of the incident wave and so depends on the experimental setup. Once A has been chosen, however, the values of B and C may be obtained by matching the $x < 0$ and $x > 0$ solutions at their meeting point, $x = 0$.

It is a mathematical result (you will see a proof next year) that the value and slope of any solution of the time-independent Schrödinger equation are continuous wherever the potential $V(x)$ is finite. In our case, since the potential is finite everywhere, the value and slope are continuous everywhere. We can use this to relate the values of B and C to the value of A.

Continuity of $\phi(x)$ implies that $\lim_{x\to 0} \phi(x)$ does not depend on whether $x \to 0$ from below or above:

$$\lim_{x\to 0^-} \phi(x) = \lim_{x\to 0^+} \phi(x) \ .$$

Using the known functional forms of $\phi(x)$ when $x < 0$ and $x > 0$, we obtain

$$\lim_{x\to 0^-} \left(Ae^{ikx} + Be^{-ikx}\right) = \lim_{x\to 0^+} Ce^{-\gamma x} \ ,$$

and hence

$$A + B \; = \; C \; .$$

Similarly, continuity of $d\phi/dx$ implies that $\lim_{x\to 0} d\phi/dx$ does not depend on whether $x \to 0$ from below or above:

$$\lim_{x\to 0^-} \frac{d\phi(x)}{dx} = \lim_{x\to 0^+} \frac{d\phi(x)}{dx} \; .$$

Using the known functional forms of $\phi(x)$ when $x < 0$ and $x > 0$, we obtain

$$\lim_{x\to 0^-} \left(ikAe^{ikx} - ikBe^{-ikx} \right) = \lim_{x\to 0^+} \left(-\gamma Ce^{-\gamma x} \right) \; ,$$

and hence

$$ik(A - B) \; = \; -\gamma C \; .$$

Solving the simultaneous equations,

$$A + B \; = \; C \; , \qquad\qquad ik(A - B) \; = \; -\gamma C \; ,$$

gives B and C in terms of A:

$$B \; = \; \frac{k - i\gamma}{k + i\gamma} A \; , \qquad\qquad C \; = \; \frac{2k}{k + i\gamma} A \; .$$

Interpreting the results

As already explained, A and B are the amplitudes of the incident and reflected waves. The intensity of the reflected wave is

$$|B^2| \; = \; \left(\frac{k - i\gamma}{k + i\gamma} A \right)^* \left(\frac{k - i\gamma}{k + i\gamma} A \right) \; = \; \left(\frac{k + i\gamma}{k - i\gamma} A^* \right) \left(\frac{k - i\gamma}{k + i\gamma} A \right) \; = |A^2|.$$

Since the intensity at point x is proportional to the probability density of finding the particle at x, this shows that every particle that strikes the barrier is eventually reflected.

More interesting is the fact that C is not equal to zero. This implies that there is a non-zero probability density,

$$|Ce^{-\gamma x}|^2 = \frac{4k^2}{k^2 + \gamma^2} |A^2| e^{-2\gamma x} \; ,$$

of finding the particle at a position x inside the barrier. The exponential decrease with x implies it is very unlikely that the particle will be found far inside the barrier, but even this is not ruled out.

Barriers of finite width

For a barrier of height U and width a, there are three regions to consider:

$$\begin{cases} \text{Region I} & x < 0 & V(x) = 0 \\ \text{Region II} & 0 < x < a & V(x) = U \\ \text{Region III} & x > a & V(x) = 0 \end{cases}$$

The wave functions in the three regions are:

$$\begin{cases} \text{Region I} & \phi(x) = Ae^{ikx} + Be^{-ikx} \\ \text{Region II} & \phi(x) = Ce^{-\gamma x} + De^{\gamma x} \\ \text{Region III} & \phi(x) = Ee^{ikx} \end{cases}$$

The (mathematically acceptable) Fe^{-ikx} term in Region III, which represents a wave incident from the right, has been omitted because we are considering an experiment in which the only incident wave is from the left.

Just as for an infinitely wide barrier, the value of A depends on the experimental setup. The other four amplitudes (B, C, D and E) can be expressed in terms of A by solving the four simultaneous equations obtained by matching the wave functions and their derivatives at both edges of the barrier, $x = 0$ and $x = a$.

If you wade through the (very complicated) mathematics, you find that $|E^2|$ is greater than zero, showing that the particle may sometimes tunnel through the classically forbidden barrier region to emerge on the other side. The intensity of the reflected wave, $|B^2|$, is correspondingly smaller than $|A^2|$.

If the barrier is wide enough, the tunnelling probability $|E^2|/|A^2|$ is roughly proportional to $e^{-2\gamma x}$ (the same exponential decay seen in the expression for the probability density inside an infinitely wide barrier). This approximation is accurate whenever the tunnelling probability is low and is used in the next section on the scanning tunnelling microscope.

The scanning tunnelling microscope

As illustrated in Fig. 7.5, a scanning tunnelling microscope (STM) consists of a sharp metallic tip positioned a few Å above a metallic surface using a very accurate piezoelectric mechanism. A voltage V is applied between the tip and the surface, causing electrons to tunnel across the intervening gap. Since the measured tunnelling current depends very sensitively (exponentially) on the size of the gap, it can be used as a very precise altimeter to “fly” the tip across the surface.

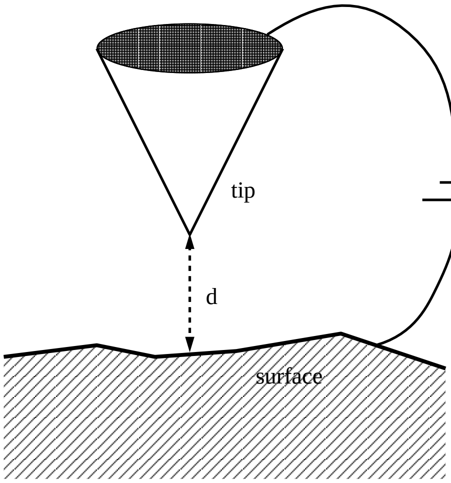


Figure 7.5: The tip of a scanning tunnelling microscope is scanned across the surface using the tunnelling current as an altimeter.

Atomic resolution is achievable because of the very high sensitivity of the tunnelling current to the tip-surface distance. It is also important that the piezoelectric positioning mechanism is precise and the vibration filtering good. Even if the tip is not ideally sharp, the rapid decay of the tunnelling probability with distance implies that only the lowest few atoms contribute to the tunnelling current. This allows atomic resolution to be achieved without using a tip that has been precisely engineered on an atomic scale. A schematic diagram of an energy eigenfunction $\phi(z)$ obtained by solving the time-independent Schrödinger equation in the tip, tunnelling barrier, and surface regions is shown in Fig. 7.6.

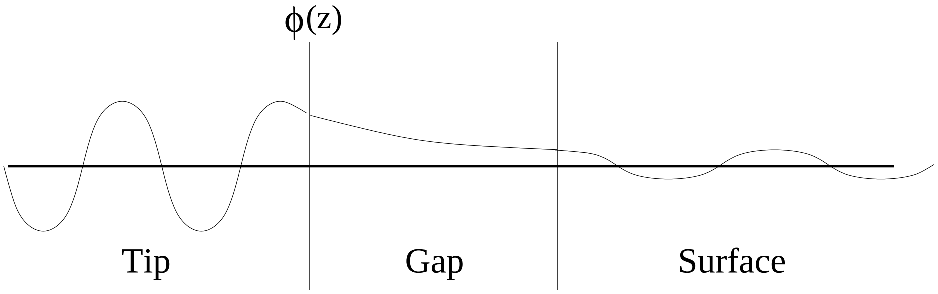


Figure 7.6: A schematic diagram of the energy eigenfunction in a scanning tunnelling microscope.

Suppose that the most energetic electrons in the metal have total energy (kinetic + potential) equal to E_F (the "Fermi energy"). Since an additional energy equal to the work function W is required to remove such an electron from the metal, the potential energy in the gap region must be $E_F + W$. Given that the applied voltage is small, we can approximate the potential in the tunnelling region as a square barrier of height W; the wave function of an electron of energy E_F attempting to tunnel from the tip to the surface then obeys the Schrödinger equation:

$$\left(-\frac{\hbar^2}{2m}\frac{d^2}{dz^2} + E_F + W\right)\phi(z) = E_F\phi(z) \; .$$

This equation has a solution of the form $e^{-\gamma z}$ with

$$\frac{\hbar^2}{2m}\gamma^2 = W \; ,$$

or, equivalently,

$$\gamma = \sqrt{\frac{2mW}{\hbar^2}} \; .$$

This gives

$$\frac{|\phi(d)|^2}{|\phi(0)|^2} = \frac{|C^2|e^{-2\gamma d}}{|C^2|} = e^{-2\gamma d} \; .$$

Since the electron densities in the tip and surface regions are proportional to the values of $|\phi|^2$ in those regions, the tunnelling probability (the ratio of electron densities) is approximately $e^{-2\gamma d}$. For a typical work function of, say, 5V, we obtain:

$$\gamma = \sqrt{\frac{2 \times 9.11 \times 10^{-31} \times 5 \times 1.60 \times 10^{-19}}{(1.05 \times 10^{-34})^2}} \approx 1.15 \times 10^{10}\,\mathrm{m}^{-1} \; .$$

The fractional change in tunnelling current when the tip-surface distance decreases by Δd is

$$\frac{e^{-2\gamma(d-\Delta d)}}{e^{-2\gamma d}} = e^{2\gamma\Delta d} \; .$$

If $\Delta d = 10^{-10}$ m (a typical atomic radius), then

$$e^{2\gamma d} = e^{2\times1.15\times10^{10}\times10^{-10}} \approx 9.97 \; .$$

The tunnelling current increases almost ten times when the tip-surface separation decreases by just one atomic radius!