1

The Principle of Wave–Particle Duality: An Overview

1.1 Introduction

In the year 1900, physics entered a period of deep crisis as a number of peculiar phenomena, for which no classical explanation was possible, began to appear one after the other, starting with the famous problem of blackbody radiation. By 1923, when the "dust had settled," it became apparent that these peculiarities had a common explanation. They revealed a novel fundamental principle of nature that was completely at odds with the framework of classical physics: the celebrated *principle of wave–particle duality*, which can be phrased as follows.

The principle of wave–particle duality**:** *All physical entities have a dual character; they are waves and particles at the same time. Everything we used to regard as being exclusively a wave has, at the same time, a corpuscular character, while everything we thought of as strictly a particle behaves also as a wave. The relations between these two classically irreconcilable points of view—particle versus wave—are*

$$E = hf, \quad p = \frac{h}{\lambda}, \tag{1.1}$$

or, equivalently,

$$f = \frac{E}{h}, \quad \lambda = \frac{h}{p}. \tag{1.2}$$

In expressions (1.1) *we start off with what we traditionally considered to be solely a wave—an electromagnetic (EM) wave, for example—and we associate its wave characteristics f and λ (frequency and wavelength) with the corpuscular characteristics E and p (energy and momentum) of the corresponding particle. Conversely, in expressions* (1.2), *we begin with what we once regarded as purely a particle—say, an electron—and we associate its corpuscular characteristics E and p with the wave characteristics f and λ of the corresponding wave. Planck's constant h, which provides the link between these two aspects of all physical entities, is equal to*

$$h = 6.62 \times 10^{-27}\,\text{erg}\,\text{s} = 6.62 \times 10^{-34}\,\text{J}\,\text{s}.$$

Actually, the aim here is not to retrace the historical process that led to this fundamental discovery, but precisely the opposite: Taking wave–particle duality as

An Introduction to Quantum Physics: A First Course for Physicists, Chemists, Materials Scientists, and Engineers, First Edition. Stefanos Trachanas, Manolis Antonoyiannakis and Leonidas Tsetseris.

granted, we aim to show how effortlessly the peculiar phenomena we mentioned earlier can be explained. Incidentally, these phenomena merit discussion not only for their historical role in the discovery of a new physical principle but also because of their continuing significance as fundamental quantum effects. Furthermore, we show that the principle of wave–particle duality should be recognized as the only sensible explanation to fundamental “mysteries” of the atomic world—such as the *extraordinary stability of its structures* (e.g., atoms and molecules) and the *uniqueness of their form*—and not as some whim of nature, which we are supposed to accept merely as an empirical fact.

From its very name, it is clear that the principle of wave–particle duality can be naturally split in two partial principles: (i) *the principle of wave–particle duality of light* and (ii) *the principle of wave–particle duality of matter*. We proceed to examine both these principles, in relation to the peculiar phenomena and problems that led to their postulation.

1.2 The Principle of Wave–Particle Duality of Light

According to the preceding discussion, the wave–particle duality says that light—which in classical physics is purely an EM wave—has also a corpuscular character. The associated particle is the celebrated quantum of light, the *photon*. The wavelike features f and λ of the corresponding EM wave, and the particle-like features E and p of the associated particle, the photon, are related through expressions (1.1). We will now see how this principle can explain two key physical phenomena—the photoelectric effect and the Compton effect—that are completely inexplicable in the context of classical physics.

1.2.1 The Photoelectric Effect

With this term we refer today to the general effect of light-induced removal of electrons from physical systems where they are bound. Such systems can be atoms and molecules—in which case we call the effect *ionization*—or a metal, in which case we have the standard photoelectric effect studied at the end of the nineteenth and the beginning of twentieth century. What makes the effect peculiar from a classical perspective is the failure of classical physics to explain the following empirical fact: The photoelectric effect (i.e., the removal of electrons) is possible only if the frequency f of the incident EM radiation is greater than (or at least equal to) a value f_0 that depends on the system from which the removal occurs (atom, molecule, metal, etc.). We thus have

$$f \geq f_0. \tag{1.3}$$

In classical physics, a “threshold condition” of the type (1.3) has no physical justification. Whatever the frequency of the incident EM wave, its electric field will always produce work on the electrons, and when this exceeds the *work function* of the metal—the minimum energy required for extraction—electrons will be ejected from it. In other words, in classical physics, the frequency plays no crucial role in the energy exchanges between light and matter, while the intensity of the electric field of light is the decisive factor. Clearly, the very existence of a threshold frequency in the photoelectric effect leaves no room for a classical explanation. In contrast, the phenomenon is easily understood in quantum mechanics. A light beam of frequency f is also a stream of photons with energy $\epsilon = hf$; therefore, when quantized light—a "rain of light quanta"—impinges on a metal, only one of two things can happen: Since the light quantum is by definition indivisible, when it "encounters" an electron it will either be absorbed by it or "pass by" without interacting with it.[1] In the first case (absorption), the outcome depends on the relative size of $\epsilon = hf$ and the work function, W, of the metal. If the energy of the light quantum (i.e., the photon) is greater than the work function, the photoelectric effect occurs; if it is lower, there is no such effect. Therefore, the quantum nature of light points naturally to the existence of a threshold frequency in the photoelectric effect, based on the condition

$$hf \geq W \Rightarrow f \geq \frac{W}{h} = f_0, \tag{1.4}$$

which also determines the value of the threshold frequency $f_0 = W/h$. For $hf > W$, the energy of the absorbed photon is only partially spent to extract the electron, while the remainder is turned into kinetic energy K $(= mv^2/2)$ of the electron. We thus have

$$hf = W + K = W + \frac{1}{2}mv^2, \tag{1.5}$$

which is known as *Einstein's photoelectric equation*. Written in the form

$$K = hf - W \qquad (f \geq f_0), \tag{1.6}$$

Equation (1.5) predicts a linear dependence of the photoelectrons' kinetic energy on the light frequency f, as represented by the straight line in Figure 1.1.

Therefore, by measuring K for various values of f we can fully confirm—or disprove—Einstein's photoelectric equation and, concomitantly, the quantum nature of light, as manifested via the photoelectric effect. In addition, we can deduce the value of Planck's constant from the slope of the experimental line.

The discussion becomes clearer if in the basic relation $\epsilon = hf = hc/\lambda$ we express energy in electron volts and length in angstroms—the "practical units" of the atomic world ($1\,\text{Å} = 10^{-10}\,\text{m}$, $1\,\text{eV} = 1.6 \times 10^{-19}\,\text{J} = 1.6 \times 10^{-12}\,\text{erg}$). The product hc, which has dimensions of energy times length (since h has dimensions of energy times time), then takes the value $hc = 12\,400\,\text{eV Å}$, and the formula for the energy of the photon is written as

$$\epsilon(\text{eV}) = \frac{12\,400}{\lambda(\text{Å})} \approx \frac{12\,000}{\lambda(\text{Å})}. \tag{1.7}$$

---

[1] For completeness, let us also mention the possibility of *scattering*. Here, the photon "collides" with an electron, transfers to it part of its energy and momentum, and scatters in another direction as a photon of different frequency (i.e., a *different photon*). This is the *Compton effect*, which we examine in the coming section. But let us note right away that Compton scattering has negligible probability to occur for low-energy photons like those used in the photoelectric effect.

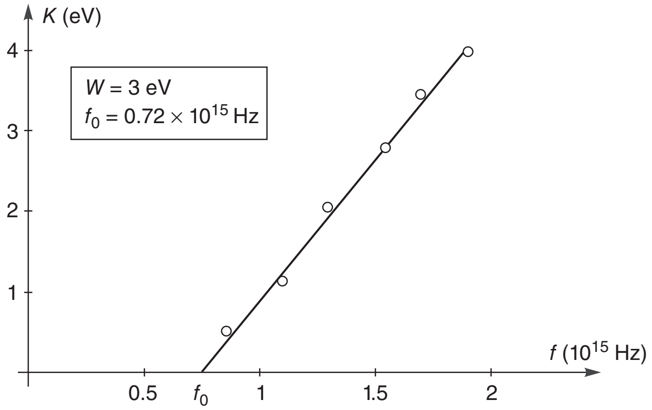


Figure 1.1 The kinetic energy K of electrons as a function of photon frequency f. The experimental curve is a straight line whose slope is equal to Planck's constant.

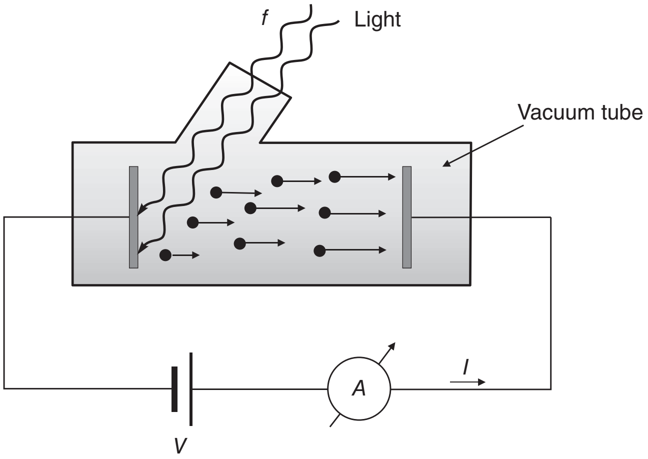


Figure 1.2 The standard experimental setup for studying the photoelectric effect. The photoelectric current occurs only when $f > f_0$ and vanishes when f gets smaller than the threshold frequency f_0. The kinetic energy of the extracted electrons is measured by reversing the polarity of the source up to a value V_0—known as the *cutoff potential*—for which the photoelectric current vanishes and we get $K = eV_0$.

The last expression is often used in this book, since it gives simple numerical results for typical wavelength values. For example, for a photon with $\lambda = 6000$ Å—at about the middle of the visible spectrum—we have $\epsilon = 2$ eV. We remind the readers that the electron volt (eV) is defined as the kinetic energy attained by an electron when it is accelerated by a potential difference of 1 V. Figure 1.2 shows a typical setup for the experimental study of the photoelectric effect. Indeed, Einstein's photoelectric equation is validated by experiment, thus confirming directly that light is quantized, as predicted by the principle of wave–particle duality of light.

Example 1.1 A beam of radiation of wavelength $\lambda = 2000\,\text{Å}$ impinges on a metal. If the work function of the metal is $W = 2\,\text{eV}$, calculate: (i) the kinetic energy K and the speed v of the photoelectrons, (ii) the cutoff potential V_0.

Solution: If we set $\lambda = 2000\,\text{Å}$ in the relation $\epsilon(\text{eV}) = 12\,000/\lambda(\text{Å})$, we obtain $\epsilon = 6\,\text{eV}$. So if we subtract the work function 2 eV, we obtain 4 eV for the kinetic energy of the outgoing electrons. The speed of the photoelectrons can then be calculated by the relation

$$\begin{aligned} K &= \frac{1}{2}mv^2 = \frac{1}{2}mc^2\left(\frac{v}{c}\right)^2 \Rightarrow \frac{v}{c} = \sqrt{\frac{2K}{mc^2}} \\ &= \sqrt{\frac{2\times 4\,\text{eV}}{\frac{1}{2}\times 10^6\,\text{eV}}} = 4\times 10^{-3} \Rightarrow \frac{v}{c} = 4\times 10^{-3} \Rightarrow v = 1.2\times 10^8\,\text{cm/s}. \end{aligned}$$

Here we wrote $\frac{1}{2}mv^2$ as $\frac{1}{2}mc^2(v/c)^2$ in order to express mc^2 in eV ($mc^2 = 0.5\,\text{MeV}$ for an electron) and the electronic speed as a fraction of the speed of light (which is useful in several ways: for example, it helps us assess the validity of our nonrelativistic treatment of the problem). As for the cutoff potential V_0, it is equal to $V_0 = 4\,\text{V}$, since $K = 4\,\text{eV}$ and $K = e \cdot V_0$.

We should pause here to remark how much simpler and more transparent our calculations become when, instead of using the macroscopic units of one system or another (cgs or SI), we use the "natural" units defined by the very phenomena we study. For example, we use eV for energy, which also comes in handy when we express the rest mass of particles in terms of their equivalent energy rather than in g or kg. In this spirit, it is worthwhile to memorize the numbers $m_ec^2 \approx 0.5\,\text{MeV}$ and $m_pc^2 \approx 1836\,m_ec^2 = 960\,\text{MeV} \approx 1\,\text{GeV}$ for electrons and protons, respectively. We will revisit the topic of units later (Section 1.2.3).

1.2.2 The Compton Effect

According to expressions (1.1), a photon carries energy $\epsilon = hf$ *and* momentum $p = h/\lambda$. And because it carries momentum, the photon can be regarded as a particle in the full sense of the term. But how can we verify that a photon has not only energy but also momentum? Clearly, we need an experiment whereby photons collide with very *light* particles—we will shortly see why. We can then apply the conservation laws of energy and momentum during the collision to check whether photons satisfy a relation of the type $p = h/\lambda$.

Why do we need the target particles to be as light as possible—that is, electrons? It is well known that when small moving spheres collide with considerably larger stationary ones, they simply recoil with no significant change in their energy, while the large spheres stay practically still during the collision. Conversely, if the target spheres are also small (or even smaller than the projectile particles), then upon collision they will move, taking some of the kinetic energy of the impinging spheres, which then scatter in various directions with lower kinetic energy. Therefore, if photons are particles in the full sense of the term, they will behave as such when scattered by light particles, like the electrons of a material: They will transfer part of their momentum and energy to the target electrons and end up with lower energy than they had before the collision. In other words, we will have

$$\epsilon' = hf' < \epsilon = hf \Rightarrow f' < f \Rightarrow \lambda' > \lambda, \tag{1.8}$$

where the primes refer to the scattered photons. This shift of the wavelength to greater values when photons collide with electrons is known as the *Compton effect*. It was confirmed experimentally by Arthur H. Compton in 1923, when an x-ray beam was scattered off by the electrons of a target material. Why were x-rays used to study the effect? (Today we actually prefer γ rays for this purpose.) Because x- (and γ) rays have very short wavelength, the momentum $p = h/\lambda$ of the impinging photons is large enough to ensure large momentum and energy transfer to the practically stationary target electrons (whereby the scattered photons suffer a great loss of momentum and energy). In a Compton experiment we measure the wavelength λ' of the scattered photon as a function of the scattering angle θ between the directions of the impinging and scattered photon. By applying the principles of energy and momentum conservation we can calculate the dependence $\lambda' = \lambda'(\theta)$ in a typical collision event such as the one depicted in Figure 1.3.

Indeed, if we use the conservation equations—see Example 1.2—to eliminate the parameters E, p, and ϕ (which we do not observe in the experiment, as they pertain to the electron), we eventually obtain

$$\Delta\lambda = \lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta) = \lambda_C(1 - \cos\theta), \tag{1.9}$$

where

$$\lambda_C = \frac{h}{mc} = 0.02427\,\text{Å} \approx 24 \times 10^{-3}\,\text{Å} \tag{1.10}$$

is the so-called *Compton wavelength* of the electron. It follows from (1.9) that the fractional shift in the wavelength, $\Delta\lambda/\lambda$, is on the order of λ_C/λ, so it is considerable in size only when λ is comparable to or smaller than the Compton wavelength. This condition is met in part for hard x-rays and in full for γ rays.

Compton's experiment fully confirmed the prediction (1.9) and, concomitantly, the relation $p = h/\lambda$ on which it was based. The wave–particle duality of light is thus an indisputable experimental fact. Light—and, more generally, EM radiation—has a wavelike and a corpuscular nature at the same time.

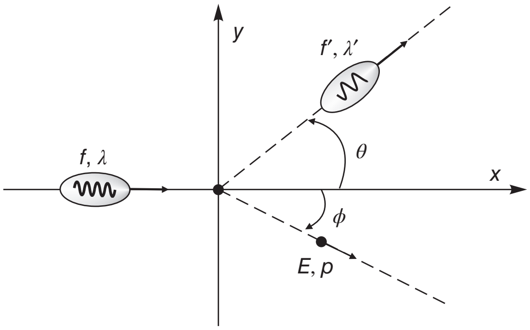


Figure 1.3 A photon colliding with a stationary electron. The photon is scattered at an angle θ with a wavelength λ' that is greater than its initial wavelength λ. The electron recoils at an angle ϕ with energy E and momentum p.

Example 1.2 In a Compton experiment the impinging photons have wavelength $\lambda = 12 \times 10^{-3}$Å $= \lambda_C/2$ and some of them are detected at an angle of 60° with respect to the direction of the incident beam. Calculate (i) the wavelength, momentum, and energy of the scattered photons and (ii) the momentum, energy, and scattering angle of the recoiling electrons. Express your results as a function of the electron mass and fundamental physical constants.

Solution: For $\lambda = \lambda_C/2$ and $\theta = 60°$ ($\Rightarrow \cos\theta = 1/2$), the formula $\Delta\lambda = \lambda' - \lambda = \lambda_C(1 - \cos\theta)$ yields $\lambda' = \lambda_C$, which is twice the initial wavelength. The momentum and energy of the photon before and after scattering are

$$p_\gamma = \frac{h}{\lambda} = \frac{h}{\lambda_C/2} = \frac{h}{(h/mc)/2} = 2mc, \quad p'_\gamma = \frac{h}{\lambda'} = \frac{h}{\lambda_C} = mc$$

and

$$\epsilon = hf = \frac{hc}{\lambda} = \frac{hc}{\lambda_C/2} = 2mc^2, \quad \epsilon' = \frac{hc}{\lambda'} = \frac{hc}{\lambda_C} = mc^2,$$

where the index "γ" in the momentum symbol p denotes the photon (in customary reference to "γ rays") to disambiguate it from the symbol p of the electronic momentum. We can now write the conservation laws of energy and momentum as follows:

- *Conservation of energy*

$$\epsilon + mc^2 = \epsilon' + E \Rightarrow 2mc^2 + mc^2 = mc^2 + E \Rightarrow E = 2mc^2.$$

- *Conservation of momentum along the x-axis* (Figure 1.3 with $\theta = 60°$)

$$p_\gamma + 0 = p'_\gamma \cos\theta + p\cos\phi \Rightarrow 2mc + 0 = mc\,\frac{1}{2} + p\cos\phi$$
$$\Rightarrow p\cos\phi = \frac{3}{2}mc. \qquad (1)$$

- *Conservation of momentum along the y-axis*

$$0 + 0 = p'_\gamma \sin\theta - p\sin\phi \Rightarrow 0 = mc\,\frac{\sqrt{3}}{2} - p\sin\phi$$
$$\Rightarrow p\sin\phi = mc\,\frac{\sqrt{3}}{2}. \qquad (2)$$

If we now take the square of (1) and (2) and add them, we get

$$p^2 = 3m^2c^2 \Rightarrow p = \sqrt{3}\,mc$$

and, based on (1), we find that $\sqrt{3}\,mc\cos\phi = (3/2)mc \Rightarrow \cos\phi = \sqrt{3}/2 \Rightarrow \phi = 30°$.

Now that p and E for the electron ($p = \sqrt{3}\,mc, E = 2mc^2$) have been calculated, one may wonder whether they satisfy the relativistic energy–momentum relation $E^2 = c^2p^2 + m^2c^4$.[2] Indeed they do, as the readers can readily verify.

---

[2] The use of relativistic formulas here is necessary because the speeds of the recoiling electrons are indeed relativistic.

1.2.3 A Note on Units

At this point we should pause to make some remarks on the system of units. We have already suggested (see Example 1.1) that both the cgs and SI system of units are equally unsuited for the atomic world, since it would be quite unreasonable to measure, for example, the energy in joules (SI) or erg (cgs). The natural scale of energies in atoms is the electron volt, a unit that is 19 orders of magnitude smaller than the joule and 12 orders smaller than the erg! Likewise, the natural length unit in the atomic world is the angstrom ($= 10^{-10}$ m), since it is the typical size of atoms. In this spirit, it is inconvenient to express, say, Planck's constant in erg s or J s, and hc—another useful constant—in erg cm or J m; instead, it is easier to use the corresponding practical units eV s for h and eV Å for hc. There is, however, one instance in atomic physics where we cannot avoid choosing one system over another: The basic force law governing atomic and molecular structure—Coulomb's law—has a much more convenient form in cgs than SI units, namely,

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}\ \text{(SI)} \qquad F = \frac{q_1 q_2}{r^2}\ \text{(cgs)}, \tag{1.11}$$

whence we immediately see why the cgs system is preferable over SI in atomic physics. In SI units all mathematical expressions of the basic quantum results for atoms appear much less elegant because they carry the cumbersome factor $1/4\pi\epsilon_0$. For example, in cgs units the quantum formula for the ionization energy of the hydrogen atom has the simple form $W_\text{I} = me^4/2\hbar^2$, while in SI units it becomes $W_\text{I} = me^4/32\pi^2\epsilon_0^2\hbar^2$! Therefore, our choice is to go with the cgs system for the mathematical expression of Coulomb's law, but to make all calculations in the practical units eV and Å, or even in the so-called *atomic system of units*, which we will introduce later. As you will soon find out, the practical unit of energy (i.e., the eV) is much better suited than the joule, even for calculations concerning physical quantities, like voltage or electric field intensities in atoms, where the SI units (V or V/m) are certainly preferable. The reason is that the energy unit eV is directly related both to the fundamental unit of charge e and the SI unit of volt. A pertinent example was the calculation—without much effort!—of the cutoff potential in Example 1.1. The same holds true for electric field intensities in atoms, where the SI unit V/m (or V/cm) arises naturally from the energy unit eV.

So, even readers who are adherents of the SI system will find that the practical energy unit, eV, is much closer to the SI system than the joule itself.

As for the cgs system, we remind the readers that its basic units—length, mass, and time—are the centimeter (cm), the gram (g), and the second (s), while for derivative quantities such as force, energy, and charge, the cgs units are the dyn, the erg, and the esu-q (electrostatic unit of charge), respectively. These units are related to their SI counterparts as follows:

Quantity	cgs	SI	Conversion
Force	dyn	N (newton)	1 N = 10^5 dyn
Energy	erg	J (joule)	1 J = 10^7 erg
Charge	esu-q	C (coulomb)	1 C = 3×10^9 esu-q

Another advantage of the cgs system is that the charge q has mechanical units (due to the form of Coulomb's law), while in the SI system charge is an independent physical quantity whose unit is not related to the mechanical units of the system. Therefore, in the SI system, dimensional analysis—which we use extensively in this book—becomes unnecessarily cumbersome, since there are now four fundamental quantities instead of three. Conversely, in the cgs system, the electric charge q—or rather, its square, via the relation $F = q^2/r^2$—has mechanical units, namely,

$$[q^2] = [F \cdot r^2] = [F \cdot r \cdot r] = E \cdot L \;\Rightarrow\; [q^2] = \text{erg cm}.$$

Incidentally, another quantity with dimensions of energy times length, like q^2, is the product hc. The ratio q^2/hc is thus a dimensionless quantity, which we shall encounter later on in this chapter.

Problems

1.1 The ionization energy of the hydrogen atom is $W_{\rm I} = 13.6\,\text{eV}$. Will there be a photoelectric effect (i.e., ionization of the atom) if it is exposed to ultraviolet (UV) light of wavelength 480Å? What is the speed of the extracted electron?

1.2 Besides the threshold frequency, another remarkable feature of the photoelectric effect is the practically vanishing time between the incidence of the light beam on the photocathode and the extraction of electrons. Even for a very weak beam, photoelectrons are produced almost instantaneously ($\tau < 10^{-9}\,\text{s}$). To see how the classical theory fails here also, estimate the time needed to extract an electron from an atom exposed to a light beam of the same intensity as, say, a light bulb of 100 W at a distance of 1 m. Treat the atom as a light collector that absorbs all EM energy incident on its cross-sectional area.

1.3 A photon of $\lambda = \lambda_{\rm C}$ impinges on a stationary electron and scatters at an angle of 180° ($\theta = \pi$). Calculate the momentum and energy of the electron after the collision and confirm your results by performing an appropriate test. What could this test be?

1.4 A photon of $\lambda = \lambda_{\rm C}/2$ is Compton scattered by an initially stationary electron. (a) Calculate (in terms of h, m, and c) the wavelength, momentum, and energy of the photon scattered at a 120° angle. (b) Calculate the scattering angle, momentum, and energy of the electron after the collision.

1.3 The Principle of Wave–Particle Duality of Matter

As emphasized in the introduction, relations (1.2) of the wave–particle duality of matter are similar to those of light—relations (1.1)—but they have to be viewed in reverse order. In case (1.2), we are talking about entities (e.g., electrons) we used to recognize as *particles* in classical physics (so they are described by their energy E and momentum p), but we now learn they are also *waves*. Their wave features f and λ are connected to the corpuscular attributes E and p via relations (1.2). The electron—the most fundamental particle of nonnuclear matter—is thus a *particle* and a *wave* at the same time. We were already aware of its corpuscular nature; after all, we first came across the electron as a particle. So we just need to examine if it is also a wave with $\lambda = h/p$, as Louis de Broglie first hypothesized in 1923. Let us examine how we can infer the existence of these waves.

1.3.1 From Frequency Quantization in Classical Waves to Energy Quantization in Matter Waves: The Most Important General Consequence of Wave–Particle Duality of Matter

To experimentally verify the wavelike nature of electrons, the obvious test is to look for *interference phenomena* between electronic waves, just as in classical waves. This would be a direct confirmation. But there is also an indirect confirmation, invoking a characteristic feature of *standing waves*, namely, *frequency quantization*. A standing classical wave—localized on a finite object—can only exist if its frequency takes a discrete sequence of values known as the *eigenfrequencies* of the system. The most representative examples are the standing waves of definite frequency—the so-called *normal modes*—on a string. As it follows from Figure 1.4, the allowed frequencies of the string's vibrations—$f = c/\lambda$, where c is the speed of wave propagation—are given by

$$L = n\frac{\lambda}{2} \Rightarrow \lambda = \frac{2L}{n} \Rightarrow f = \frac{c}{\lambda} = \frac{c}{(2L/n)} = \frac{c}{2L} \cdot n, \tag{1.12}$$

which means that the only possible vibrations of the string are those with integer multiples of the fundamental frequency $f_1 = c/2L$.

But if the frequency is quantized in classical systems, so too will be the particle's energy, since the wave–particle duality of particles—namely, the relation $E = hf$—provides a direct link between their energy and the frequency of the corresponding wave. So if a quantum particle, say an electron, is trapped somewhere in space (e.g., in an atom or a molecule), the associated de Broglie wave will be a standing wave with quantized frequency, and therefore the energy $E = hf$ of the electron will also be quantized. As we will see shortly, energy quantization for particles that are trapped in some region of space (and thus perform confined motion) is the deepest consequence of the wave–particle duality of matter.

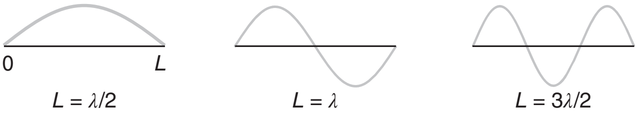


Figure 1.4 Standing classical waves on a string. A standing wave of this kind can only be formed when an integer number of half-waves fit on the string. That is, $L = n\frac{\lambda}{2}$ ($n = 1, 2, \ldots$).

1.3.2 The Problem of Atomic Stability under Collisions

We will now see that energy quantization of atomic electrons provides the only reasonable explanation for the mystery of atomic stability. Why is this a mystery? Because atoms remain completely unscathed, even though they continuously undergo violent collisions with each other. If we picture the electrons in atoms orbiting around the nucleus, like planets around a sun, then it is as if their orbits do not change at all upon innumerable collisions with other "solar systems." But there is more to it. Even if we took apart an atom—by removing all its electrons—and let it "reconstruct" itself, it would reemerge in identical form and shape. The evidence for these statements is that atoms always emit the same characteristic frequencies—the same spectrum—while their chemical behavior also remains unaltered. In fact, chemical stability is an essential prerequisite for our very existence. Note, however, that in the discussion we only considered atomic stability against collisions. We have completely ignored the stability of atoms against the radiation emitted by their electrons, which, being charged particles in accelerated motion, ought to radiate and lose energy until they fall into the nucleus. We discuss this problem in Section 1.3.6. Until then, let us simply accept that, for some reason, the classical laws of EM radiation do not hold in the atomic world. Let us then see how the problem of atomic stability against collisions can be explained naturally by assuming that the energy of atomic electrons is quantized. In the hydrogen atom, for example, if the electron, which has quantized energy, occupies the lowest possible state—the so-called *ground state*—then the smallest possible change for the atom is a transfer of the electron from the ground state to the next available state, namely, the *first excited state*. In other words, the electron can only make a discontinuous transition—a *quantum jump* (or *leap*)—toward an excited atomic state. Now, if the environment offers less energy to the atom than the energy required for such a leap, as is the case for thermal collisions at room temperature, then no transition can occur at all. Indeed, the energy difference between the ground and first excited states of any atom—or molecule—is on the order of a few eV, while the average thermal energy at room temperature is about a 100th of this value. As a result, thermal collisions at room temperature do not provide sufficient energy to excite the atoms, which thus behave as stable and invariant entities during collisions. We have to reject the classically allowed small, gradual changes in energy, and consider only those quantum jumps for which the minimum required energy is available. Hence, energy quantization arises as the only conceivable explanation of the mystery of atomic stability. The "equation"

Quantization = stability

emerges thus as the fundamental *conceptual equation* of quantum physics. And since the only natural mechanism of quantization we are aware of involves standing waves, the following reasoning also applies:

Stability → quantization → wavelike behavior.

This explanation of the central mystery of the atomic world—the remarkable stability of its structures—demonstrates that the notion of wavelike behavior for particles is not so "crazy" after all. In hindsight, we can regard it as the only natural explanation of the most fundamental problem put forward by the study of matter at the atomic level.

Example 1.3 To appreciate the extreme conditions under which atoms manage to retain their structural stability, calculate the typical frequency of collisions between air molecules. Treat air particles as spheres of 1 Å diameter and assume an approximate value for the particle density n of air equal to 10^{20}molecules/cm^3, which is about 1000 times less than the density of solid matter.

Solution: Our approach is to first estimate the *mean free path* ℓ of air molecules—that is, the mean path traversed by a particle between collisions—and then divide it by the mean thermal speed v to obtain the average time τ between two consecutive collisions. The frequency of collisions will then be equal to $1/\tau$. We can easily calculate the mean free path if we realize that a molecule collides with another when it travels far enough to cover the whole volume available to it. This volume is equal to $V/N = 1/(N/V) = 1/n$, where n is the number of particles per cm^3. Therefore, the quantities ℓ (mean free path), σ (cross section of molecules), and n (particle density) are all related via the expression

$$\begin{aligned}\ell \cdot \sigma &= \text{volume covered by a molecule that travels a distance } \ell \\ &= \text{volume of the space available per molecule} = 1/n \\ &\Rightarrow \quad \ell = \frac{1}{n\sigma}. \qquad (1)\end{aligned}$$

For $n \approx 10^{20}\,\text{cm}^{-3}$ and $\sigma \approx (10^{-8}\,\text{cm})^2 = 10^{-16}\,\text{cm}^2$, we find

$$\ell \approx 10^{-4}\,\text{cm}. \qquad (2)$$

We can also obtain the mean thermal speed v—actually, the *root-mean-square* (rms) *speed*—of the air molecules (of mass M) as follows:

$$\begin{aligned}\frac{1}{2}Mv^2 = \frac{3}{2}kT \Rightarrow Mc^2\left(\frac{v}{c}\right)^2 = 3kT \Rightarrow \frac{v}{c} = \sqrt{\frac{3kT}{Mc^2}} \\ \Rightarrow \frac{v}{c} \approx \sqrt{\frac{3 \cdot \frac{1}{40}\,\text{eV}}{30 \times 10^9\,\text{eV}}} \Rightarrow v \approx 10^5\,\text{cm/s}. \qquad (3)\end{aligned}$$

Here, we expressed the kinetic energy as $\frac{1}{2}Mc^2(v/c)^2$ in order to use familiar numbers such as the proton's rest energy $m_pc^2 \approx 2000 \cdot m_ec^2 \approx 2000 \cdot 0.5\,\text{MeV} \approx 1\,\text{GeV} = 10^9\,\text{eV}$. As a typical molecule of air we take the nitrogen molecule N_2 with mass 28 ($\approx$ 30) times that of a proton. For kT at room temperature we used the rounded value $kT \approx 1/40\,\text{eV}$, which results from $(kT)_{T\approx 12\,000\,\text{K}} \approx 1$ eV. Naturally, we have rounded the numbers significantly, as we are only making an order-of-magnitude estimate.

From expressions (2) and (3), we obtain the time between collisions $\tau = \ell/v \approx 10^{-9}$ s, so the frequency of collisions is $f = 1/\tau \approx 10^9\ \text{s}^{-1}$. Each air molecule thus undergoes approximately one billion collisions with other molecules per second. And yet it remains intact. Surely, molecules are very robust structures!

1.3.3 The Problem of Energy Scales: Why Are Atomic Energies on the Order of eV, While Nuclear Energies Are on the Order of MeV?

The main idea of the previous discussion—namely, that microscopic particles in confined motion inside a structure (such as an atom or a molecule) are represented by standing matter waves—helps us understand another central mystery of the atomic world: The smaller the region inside which a particle resides, the greater the energy of that particle. The most typical examples of this mystery are the *atom* and the *nucleus*. Atomic electrons (of the outer shell, for heavy atoms) have energies on the order of a few eV, while the corresponding energies for protons and neutrons inside the nucleus are one million times greater—that is, on the order of a few MeV! Again, the explanation lies in the wave–particle duality expression $\lambda = h/p$ and the realization that the first (fundamental) standing wave in a region of space—recall the example of the string—has a wavelength λ on the order of the linear size of the region. The wavelengths of the higher standing waves are even smaller. So we can say that the largest wavelength—the one that corresponds to the ground state—will be about the size of

$$\lambda_{\max} \approx 2L,$$

where L is the linear size of the region within which the standing wave is formed. In this case, the relation $\lambda = h/p \Rightarrow p = h/\lambda$ shows that the momentum of the trapped wave–particle cannot be smaller than

$$p_{\min} = \frac{h}{\lambda_{\max}} \approx \frac{h}{2L}.$$

And, if we are interested in the state of lowest energy—which is certainly the most important state—then $p \approx p_{\min}$, and the formula

$$p \approx \frac{h}{2L}$$

provides a good *estimate* of the momenta of particles trapped inside a quantum system of linear dimension L. For the corresponding kinetic energy, $p^2/2m$, of these particles, we have

$$K \approx \frac{h^2}{8mL^2}. \tag{1.13}$$

The conclusion is now clear: The smaller the region inside which a particle is moving, the smaller the wavelength (in the first standing wave) of that particle and, consequently, the greater its momentum and energy. Figure 1.5 should help visualize the physics of this key fact.

If we now apply formula (1.13) on a nucleon—where $m = m_{\mathrm{N}}$ =mass of a proton or neutron and $L \approx 2R$ (R =nuclear radius)—or an electron of an outer atomic shell—where $m = m_e$ and $L = 2a$ (a =atomic radius)—we obtain

$$K_{\mathrm{N}} = \frac{h^2}{32 m_{\mathrm{N}} R^2} = \frac{h^2}{32 m_e a^2} \frac{a^2}{R^2} \frac{m_e}{m_{\mathrm{N}}} = K_e \left(\frac{a}{R}\right)^2 \frac{m_e}{m_{\mathrm{N}}}. \tag{1.14}$$

Given now that $a \approx 1\,\text{Å} \approx 10^{-10}$ m, $R \approx 10^{-15}$ m, and $m_{\mathrm{N}} \approx 1836\, m_e$, expression (1.14) yields

$$K_{\mathrm{N}} \approx (10^6 - 10^7) K_e, \tag{1.15}$$

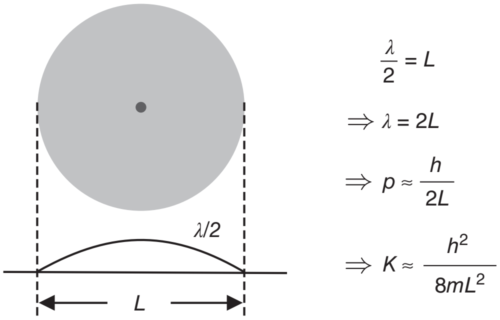


Figure 1.5 A standing matter wave of spherical shape. A particle trapped inside a bounded region—a spherical volume in our case—of linear dimension *L*, is described (in the state of lowest energy) by a spherical standing wave that vanishes only at the boundary of this region. For its wavelength we thus have $\lambda/2 \approx L \Rightarrow \lambda \approx 2L$.

which tells us that the kinetic energies K_{N} of protons and neutrons inside the nucleus are a few million times greater than the kinetic energies K_e of the outer-shell electrons in atoms. (Inner-shell electrons have greater energies than electrons in the outer shells, since they are moving in a smaller region of space.)

Furthermore, we can use the formula $K_e = h^2/32m_ea^2$ to obtain a typical value for the kinetic energy of the outer electrons

$$K_e \approx \text{a few eV}. \tag{1.16}$$

We combine Eq. (1.15) with Eq. (1.16) to obtain

$$K_{\mathrm{N}} \approx \text{a few MeV}. \tag{1.17}$$

If we now take the next logical step, namely, that the energies released in chemical and nuclear reactions should be on the same order of magnitude as the energies of outer-shell atomic electrons and nucleons, respectively, then we can deduce another fundamental feature of our world: Energies released in chemical reactions can only be on the order of an eV, while energies released in nuclear reactions must be on the order of an MeV per reaction. We can thus say that eV and MeV define the *chemical* and *nuclear energy scales*, respectively.

We can now reexamine the problem of atomic stability. If the energy scale of electrons in atoms—in the hydrogen atom, for simplicity—is on the order of an eV, then differences between adjacent energy levels (remember, they are *quantized*) should be of the same order, that is, a few eV. Note, for example, that the first excited state of the atom will correspond to a standing matter wave with $\lambda = L$ (one-half that of the ground state), so the electronic momentum $p = h/\lambda$ will double and the kinetic energy will quadruple compared to the ground state. (Provided, of course, that all standing waves of the atom occur within the same volume in space, which is not exactly true; in its excited states the atom is bigger.) The energy difference between the ground and first excited states of an atom, such as hydrogen, would thus also be on the order of an eV; this energy difference determines the atom's stability against collisions, as we noted earlier. We remind the readers that the factor kT that determines the average magnitude of thermal energies at temperature T via the relation

$$\overline{K} = \frac{3}{2}kT \tag{1.18}$$

takes (at room temperature) the approximate value

$$kT|_{T\approx 300\,\mathrm{K}} \approx \frac{1}{40}\,\mathrm{eV}. \tag{1.19}$$

We can thus see that thermal collisions at room temperature—but also at much higher temperatures, say, a few thousand degrees—cannot cause atomic excitations. Atoms emerge from their incessant collisions—roughly one billion collisions per second, as we saw—completely intact. In reality, not all atoms of a gas have the same thermal kinetic energy—(1.18) is merely a mean value—but obey a Maxwell–Boltzmann distribution, so some of them are much more energetic than others and able to cause mutual excitations when they collide with each other. So the exact picture is this: Even at room temperature, a small fraction of atoms in a gas are excited, but the overwhelming majority remains intact in their ground state.

In the case of an atomic nucleus, where the energy difference between the ground and first excited states is on the order of an MeV, a similar reasoning leads us to conclude that nuclear stability against collisions is a million times greater than atomic stability. The critical temperature for the stability of a nucleus is thus a few billion degrees kelvin, compared to a few thousand degrees for an atom. Therefore, for thermonuclear reactions to occur, as in the interior of a star, the temperature needs to rise to billions of degrees! And yet thermonuclear reactions inside stars occur—for without such reactions, we would not exist—even though the typical temperature in their interior is no greater than 10–20 million degrees! The resolution of this mystery has a quantum origin also and is discussed in the *online supplement* of Chapter 5.

1.3.4 The Stability of Atoms and Molecules Against External Electromagnetic Radiation

There are *two* types of external “perturbations” that atoms and molecules are often subjected to, and which could threaten their structural stability. The first perturbation is *thermal collisions*—actually, electric forces between electrons of approaching atoms—which we have already examined. The second type of perturbation is the ubiquitous *electromagnetic radiation*—visible light, infrared (IR), UV, x-rays, radio waves, and so on—that hits atoms continuously. Does EM radiation change the structure of atoms? If the atoms were classical systems, then the answer would surely be in the affirmative, since they would have to “respond” to any external perturbation, however small, by changing their structure accordingly; for example, by slightly adjusting their electronic orbits. However, atoms are not classical but quantum systems and therefore their states have quantized energies that can only change via specific quantum jumps. In other words, atoms cannot absorb an arbitrary amount of energy, but only the amount required for a transition from the ground state—if this is where they start from—to any one of their excited states. Now, due to the wave–particle duality of light, the incident EM radiation on an atom is also quantized with an energy quantum equal to

$$\epsilon = hf = \frac{hc}{\lambda} \Rightarrow \epsilon(\mathrm{eV}) \approx \frac{12\,000}{\lambda(\mathrm{\AA})}. \tag{1.20}$$

For example, for *visible light*, where

$$4000\,\text{Å} < \lambda < 7400\,\text{Å} \quad \text{(visible light)}, \tag{1.21}$$

the energies of optical photons span the range

$$1.6\,\text{eV} < \epsilon < 3\,\text{eV} \quad \text{(visible light)} \tag{1.22}$$

with a typical value—for $\lambda \approx 6000\,\text{Å}$—equal to 2 eV. Thus, UV photons—being more energetic, and hence more chemically potent, than optical photons—have energies greater than 3 eV, while IR photons have energies less than 1.6 eV. In other words, UV- and IR-light photon energies lie to the right and left, respectively, of the visible range (1.22). For radiowaves—where $f \approx 100\,\text{MHz} \Rightarrow \lambda = c/f = 3\,\text{m} = 3 \times 10^{10}\,\text{Å}$—we have

$$\epsilon\,(\text{radiowaves}) = \frac{12\,000}{3 \times 10^{10}}\,\text{eV} = 0.4 \times 10^{-6}\,\text{eV} \approx 1\,\mu\text{eV}.$$

What happens when one of the aforementioned kinds of radiation impinges on an atom? Take, for example, the hydrogen atom, for which the first excitation energy—equal to the energy difference between its first excited and ground states—is 10.2 eV. Clearly, any radiation whose photons have energies less than 10.2 eV cannot induce any changes to the hydrogen atom. The photons of the impinging radiation "bounce off" the atom to another direction; they are scattered, as we say, leaving the atom intact. Hence, we can conclude that energy quantization of atomic electrons—and the corresponding energy scale on the order of an eV—combined with light quantization, ensures atomic stability against not only collisions but also all of EM radiation with energy below the UV: visible, IR, microwaves, radio waves, and so on. No matter how long atoms or molecules are bombarded by such radiation—provided its intensity is not too high—they remain completely unaffected. Similarly, radiation at such frequencies cannot cause chemical reactions. The reason is that for chemical reactions there is also an energy threshold, a minimum energy barrier the light quantum has to surpass for a reaction to occur. And just like the typical energies in atoms and (small) molecules are on the order of a few eV, this threshold energy is also on the order of a few eV—typically greater than 3 eV. So the only kinds of radiation that are chemically potent are those in the UV range and beyond (x-rays, γ rays, etc.). This means, among other things, that visible light is not chemically dangerous—for if it were, we would not be here (since our planet is awash with it)!

We thus come to the realization that the crucial feature of the photoelectric effect—the existence of a threshold frequency (and energy) for the phenomenon to occur—is completely general: It holds for chemical reactions, excitations, dissociations of molecules, and so on. As a consequence, all radiation with photon energies below the energy threshold is "harmless," in the sense that it cannot cause the abovementioned effects. Given also that all "threshold" energies are on the order of a few eV, atoms and molecules are completely "safe" against all radiation from the visible range and below (in energy). (Visible light can actually cause some reactions, but these belong to a very specific category.)

1.3.5 The Problem of Length Scales: Why Are Atomic Sizes on the Order of Angstroms, While Nuclear Sizes Are on the Order of Fermis?

So far, we considered the problem of energy scales in the microscopic world. We wondered if there is a simple explanation to a seemingly paradoxical feature of nature: The energies trapped in a nucleus—in spite of its minuscule size—are a million times greater than in an atom, even though the latter is roughly 100 000 times larger. We saw that a nucleus is an energy giant precisely because it is so small. Let us recall the reason for this: Since the "fundamental" standing wave in a region of space of linear size L has a wavelength on the order of L ($\lambda \sim 2L$), the momentum $p = h/\lambda$ of the particle trapped there will be inversely proportional to L and will thus increase as the region shrinks in size. So the fact that the nucleus is an energy giant should no longer surprise us, but should instead be viewed as a direct consequence of its size in conjunction with the principle of wave–particle duality. The tinier a structure, the more energetic the particles that lie inside it. But to determine the specific energy scales—eV in atoms and MeV in nuclei—we also have to know the length scales of these structures, namely, that they are on the order of an angstrom ($= 10^{-10}$ m) for atoms, and a fermi ($= 10^{-15}$ m) for nuclei. We took these characteristic lengths as given. So our next question is: Can we explain the characteristic length scales of atoms and nuclei? Why should the size of atoms, for example, be on the order of an angstrom and not much smaller or much larger? Let us try to answer this question, starting with the self-evident idea that the ground state of an atom—which is essentially what we are after—has to be such that the total energy (kinetic + potential) is minimized. We will take the hydrogen atom as a representative example and think first of the qualitative mechanism that possibly determines its size. There are two energy terms: potential and kinetic. The former favors a short distance between the electron and the nucleus, ideally with the electron right on the nucleus and at rest. In that case, the total classical energy[3]

$$E = K + V = \frac{1}{2}mv^2 - \frac{e^2}{r} \tag{1.23}$$

tends to minus infinity for $r = 0$ and is thus clearly minimized. But the possibility of an electron *at rest* on the nucleus exists only in classical physics. In the context of the wave–particle duality of matter, it is not possible to have a particle at rest at a specific point. In fact, the exact opposite is true: If the electron is "squeezed" in such a tiny region like that of the nucleus, or even smaller, it will develop a huge momentum—on the order of $h/4R$, where R is the nuclear radius—and a corresponding kinetic energy $K = h^2/32mR^2$ that will hurl it away from the nucleus. We thus see that the notion of an electron confined in the vicinity of the nucleus does not minimize the total energy of the atom. While the potential energy is then minimized, the kinetic energy grows with no bound. To obtain the "correct" size of the atom, we need to find its radius a for which the total energy is minimized. For the kinetic energy, we use the quantum expression

---

[3] Recall that we are using the cgs system precisely because the expression for the electrostatic potential energy, $V = q_1q_2/r$, does not contain the cumbersome factor $1/4\pi\epsilon_0$. Note also that here q_1 = proton charge = e and q_2 = electron charge = $-e$.

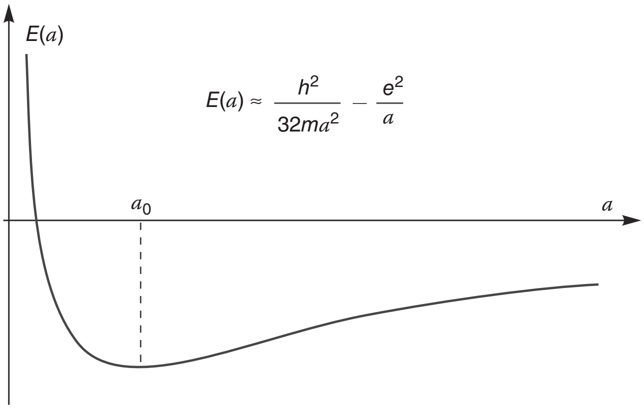


Figure 1.6 Total energy of the hydrogen atom as a function of its size, *a*. The real radius of the atom is the one that minimizes its total energy.

$K = h^2/32ma^2$, while for the potential energy we can use the approximation $V \approx -e^2/a$, even though the electron is not located at this exact distance a, since it is now a wave that extends throughout the spherical volume of radius a. (This is the so-called probability cloud as we will shortly see.) But for crude, order-of-magnitude estimates we can still use the (approximate) expression of the total energy

$$E \approx \frac{h^2}{32ma^2} - \frac{e^2}{a}. \tag{1.24}$$

Figure 1.6 shows the total energy of the atom as a function of its possible size a. The function has a minimum—obtained from the condition $dE/da = 0$—at

$$a_0 = \frac{h^2}{16me^2} \approx 1\,\text{Å}, \tag{1.25}$$

which indeed corresponds to the correct order of magnitude of atomic radii.

The general conclusion about the mechanism that determines the characteristic length scales of various atomic-scale structures is now clear. No matter how strong the mutual attraction between particles that form a microscopic structure, it will never be able to compress them to an infinitesimally small volume, because, in that case, the particles would develop an exceedingly large kinetic energy (due to their extreme localization) that would offset the energy gained by the reduction of their potential energy. The minimum total energy is thus achieved at an optimum size, which is determined by the balance between the attractive potential energy term (that pulls particles together) and the repulsive kinetic energy term (that resists their extreme localization). In this regard, the much smaller size of the nucleus compared to the atom must be accounted for by the much stronger nuclear forces exerted between the nucleons, and also by their bigger masses that weaken their resistance to localization. In fact, if we consider that the strength of nuclear forces—as measured by the relevant coupling constant g, which is the analog of e in (1.25)—is about 100 times greater than that of the EM force,[4] and also that the nucleons' mass is about 2000 times greater than an electron's mass, then formula (1.25), applied to the nucleus, gives a nuclear radius five orders of magnitude smaller than the atomic radius. This is in full agreement with observation.

Let us also add here that all atoms have roughly the same size—on the order of an Å—because their outer electrons are subject to (approximately) the same electric attraction from the nucleus as the single electron in the hydrogen atom. The reason is that inner electrons screen a large fraction of the nuclear charge from the outer electrons. In other words, from the vantage point of the outer electrons, heavier atoms resemble hydrogen and must therefore have roughly the same size.

1.3.6 The Stability of Atoms Against Their Own Radiation: Probabilistic Interpretation of Matter Waves

In 1911, the Rutherford experiment showed that the atom consists of a tiny nucleus with the electrons orbiting far away, like planets around the sun. No classical model could explain how such an atom may last more than a few tenths of a nanosecond! Whatever the classical orbit of the electrons, their motion would surely be an accelerated one (with linear or centripetal acceleration). As a result, electrons would emit EM radiation continuously, lose energy, and ultimately fall—in an infinitesimally small amount of time—on the nucleus. Conclusion: A truly *classical atom* cannot exist. But can the *quantum atom*—based on the principle of wave–particle duality of electrons—also solve the mystery of the atoms' stability against their own radiation, as it solved the two previous mysteries (stability against collisions and stability against external radiation)? Here the answer is not a resounding *yes*, as it was for the other two questions. At this point, the mystery of the stability of atoms against their own radiation cannot be solved directly, because the quantum theory has not been "set up" yet. Nevertheless, this problem can be at least bypassed with the following reasoning: If the principle of wave–particle duality is correct, then the orbital motion of electrons (which is where radiation comes from) has no physical meaning. Let us elaborate: Orbital motion means that the electron at any given time is found at a specific location (i.e., localized in space), whereas the very concept of a wave postulates a physical entity that is spread out in space. Moreover, a particle is by definition "indivisible"—within certain limits—while a wave can always be divided, for example, by letting a part of it be transmitted through one slit and another part through another. A wave is thus always *extended* and *divisible*; a particle is always *localized* and *indivisible*. At this point, the reader would be justified to ask: How can the principle of wave–particle duality of matter then be true? How can we say that a particle is at the same time a wave? How can we fit within the same physical entity two mutually exclusive properties, such as "localized and indivisible" on the one hand, and "extended and divisible" on the other? We have just arrived at the most critical question of quantum theory—a question that, as we see in the

---

[4] Measured in dimensionless units, the strength of (strong) nuclear interactions, $\alpha_s = g^2/\hbar c$, is of order 1, while the corresponding parameter $\alpha = e^2/\hbar c$ for electromagnetic forces has the known value of 1/137.

next chapter, leads to the celebrated *statistical* (or *probabilistic*) *interpretation* of matter waves. Here is what this interpretation says (Max Born, 1927):

> The function $\psi = \psi(\boldsymbol{r})$ that describes a matter wave (its so-called *wavefunction*) does not represent a measurable physical quantity. It is rather a mathematical wave—a probability wave—whose squared amplitude $|\psi(\boldsymbol{r})|^2$ yields the probability per unit volume to locate the particle in the vicinity of an arbitrary point $\boldsymbol{r}$.

We thus have

$$P(\boldsymbol{r}) = |\psi(\boldsymbol{r})|^2, \tag{1.26}$$

where $P(\boldsymbol{r})$ is the probability per unit volume—the *probability density*—of locating the particle in the vicinity of an arbitrary point in space. The total probability of finding the particle anywhere in space is given by the integral over all space

$$\int |\psi(\boldsymbol{r})|^2 \, dV = 1, \tag{1.27}$$

which clearly equals unity.

Given this interpretation, the wavefunction ψ has no immediate physical meaning—as it does not represent some sort of a physical wave—so it can take complex values in general. This is why absolute values are necessary in (1.26) or (1.27). According to (1.26), quantum particles frequent locations where their wave is strong—“stormy” areas—and avoid “calm” places where their wave is weak. In the context of such an interpretation, the contradiction between particles and waves is removed at once, since the particle need not cease being a particle and does not have to physically “disperse” throughout the volume of the wave. The wave simply describes the probability of detecting the particle here *or* there, but never here *and* there at the same time. When we do locate the particle, our detectors always record an integral and indivisible entity. No experiment has ever “captured” half an electron or a quarter of a proton. To give readers an idea of how we describe quantum particles, we depict in Figure 1.7 two simple examples of one-dimensional wavefunctions.

This short detour in our discussion helped us arrive at the following basic conclusion: The correct interpretation of the principle of wave–particle duality strips the concept of electronic orbits in atoms of any physical meaning. As a result, it makes no sense to speak of accelerated motion of electrons, nor, therefore, of emission of radiation from them. In other words, we do not have a solution to the problem mentioned earlier—but we do not have a problem either! However, a new question pops up naturally at this point: How do atoms radiate after all? We address this question in the following section.

1.3.7 How Do Atoms Radiate after All? Quantum Jumps from Higher to Lower Energy States and Atomic Spectra

Let us make some “impromptu” thoughts on this question, using again the hydrogen atom as an example. Like any standing wave, a standing electron

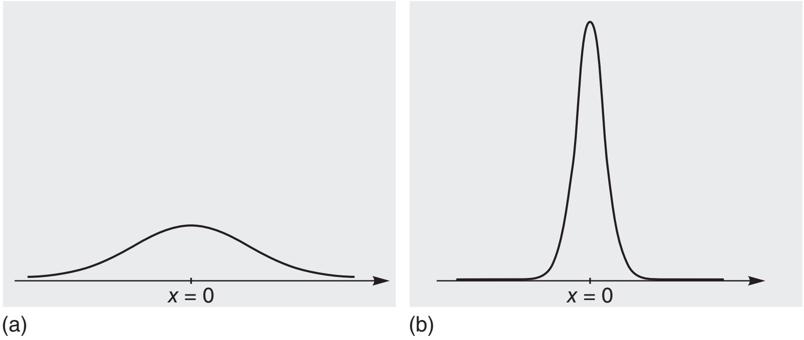


Figure 1.7 Typical one-dimensional wavefunctions. (a) An extended wavefunction: The position of the particle is known with very low precision. There is a significant probability of locating the particle in regions away from the "most frequented" location at $x = 0$. (b) A localized wavefunction: The position of the particle is known with very high precision. In the vast majority of the measurements, we would detect the particle in the immediate vicinity of $x = 0$.

wave around the nucleus can exist in a number of possible forms—the so-called *normal modes*. The first form corresponds to the state of lowest energy and the next ones correspond to excited atomic states. The corresponding energies are quantized according to some discrete sequence $E_1, E_2, \ldots, E_n, \ldots$ Since these successive standing waves around the nucleus represent the only possible energy states of the electron, we have the following two scenarios:

(a) If the electron is in the ground state, then it obviously cannot radiate; for if it did, then it would lose energy and would have to move to a lower energy state, which, however, does not exist.
(b) If the electron is in an excited state—say, the first excited state—it can be de-excited,[5] but only according to the basic quantum rules described earlier. First of all, a gradual de-excitation is impossible because the electron would then be able to gradually shed its excess energy in the form of radiation and transit to states with gradually decreasing energy, which, again, do not exist. The only available state to go to is the ground state, which, however, is located (in the hydrogen atom) 10.2 eV below the first excited state. So what can the excited electron do to "shed" its excess energy and return to the ground state? Very simply, a *quantum jump*: an abrupt transition from the excited to its ground state via emission of the energy difference 10.2 eV in the form of a UV photon.

Atoms, therefore—and, likewise, molecules, and all other quantum systems—emit light only when they undergo a transition from one of their excited states to a lower state. When this happens, a photon is emitted with energy hf, equal to the energy difference between the initial and final states of the transition. We thus have

$$E_n - E_m = hf_{nm} \qquad (n > m), \tag{1.28}$$

---

[5] As we shall later see (e.g., Chapters 9 and 16), excited states are always *unstable* and get de-excited by emission of electromagnetic radiation from the atom.

where E_n $(n > 1)$ is the energy of the initial excited state of the atom and E_m is the energy of the final state (which may or may not be its ground state).

The frequencies f_{nm} are what we observe in the so-called *line emission or absorption spectrum* of a gas made of the atoms or molecules we wish to study. We thus realize that the quantization of electronic energies in atoms or molecules is reflected in the line spectra of the corresponding substances in gas form. In turn, these spectra are our best "tool" for measuring the allowed energies in a quantum system.

The frequencies f_{nm} that correspond to the transitions $n \to m$ are known as *Bohr frequencies*. Theoretical physicists, however, prefer to use the same term for the corresponding angular frequencies, $\omega_{nm} = 2\pi f_{nm}$, because ω is better suited than f $(= \omega/2\pi)$ for the mathematical description of harmonic oscillations or waves. Note, for example, that the mathematical expression of a harmonic oscillation $x(t) = A\ \sin(2\pi t/T)$—where T is the period—takes the much simpler form $x(t) = A\ \sin \omega t$ if we introduce the angular frequency ω, via the relation

$$\omega = \frac{2\pi}{T} = 2\pi f. \tag{1.29}$$

In the same spirit, theoretical physicists prefer to write the fundamental relation $\epsilon = hf$ in the equivalent form

$$\epsilon = hf = h\frac{\omega}{2\pi} = \frac{h}{2\pi}\omega = \hbar\omega,$$

where

$$\hbar = \frac{h}{2\pi} \tag{1.30}$$

is the so-called *reduced Planck's constant*. As we will see later, the mathematical expressions of basic quantum results are considerably simplified when written in terms of $\hbar$ instead of h. Thus, the use of $\hbar$ instead of h is now common in quantum physics, while one can always revert to the older symbol whenever there is a need to use quantities closer to what is experimentally measured, such as the frequency f, or the wavelength λ.

Having thus opted to use $\hbar$ over h, we can rewrite the second expression—$p = h/\lambda$—of the wave–particle duality as

$$p = \frac{h}{\lambda} = \frac{2\pi\hbar}{\lambda} = \hbar\frac{2\pi}{\lambda} = \hbar k, \tag{1.31}$$

where

$$k = \frac{2\pi}{\lambda} \tag{1.32}$$

is the so-called *wavenumber* of the wave. Clearly, k is the spatial equivalent of ω, with λ in place of T, as we should have expected, since λ is the *spatial* and T is the *temporal period* of a sinusoidal wave. The modern version of the wave–particle duality is thus written as

$$E = \hbar\omega, \qquad p = \hbar k, \tag{1.33}$$

which is clearly more elegant than the older form.

1.3.8 Quantized Energies and Atomic Spectra: The Case of Hydrogen

With hydrogen being the simplest and most abundant element, it is no accident that the spectrum of its atoms in gaseous form has been studied extensively in the visible and its neighboring regions (IR and UV) of the EM spectrum. A remarkable result of those studies is the famous Balmer's formula for the emitted frequencies

$$f_{nm} = R\left(\frac{1}{m^2} - \frac{1}{n^2}\right), \tag{1.34}$$

where n and m are positive integers ($n > m$) and $R = 3.27 \times 10^{15}\,\mathrm{s}^{-1}$ is the so-called *Rydberg constant*, which has dimensions of frequency.

A comparison of formulas (1.28) and (1.34) gives the following expression for the quantized energies of the hydrogen atom:

$$E_n = -\frac{hR}{n^2} = -\frac{13.6}{n^2}\,\mathrm{eV}, \tag{1.35}$$

where the negative sign appears because we are talking about *bound states*. Here, the number 13.6 eV is simply the numerical value of hR expressed in units of eV.

According to (1.35), the ground-state energy of the atom is $E_1 = -13.6$ eV. Its opposite, $W_\mathrm{I} = 13.6$ eV, is the *ionization energy*, as confirmed by chemical data.

It also follows from (1.35) that the energy of the first excited state of the atom is $E_2 = -3.4$ eV, and, therefore, the first excitation energy $E_2 - E_1$ is indeed equal to the value 10.2 eV we already mentioned. This unusually large value explains the remarkable stability of the hydrogen atom against external influences (thermal collisions or EM radiation).

The conclusion from the preceding discussion should be clear. Spectral data for the hydrogen atom fully confirm all the general atomic properties we identified earlier based solely on the principle of wave–particle duality for electrons. The electronic energies in the atom are indeed quantized—they only take the discrete values of (1.35)—and the energy scale is indeed on the order of a few eV, as we predicted. And given that the energy scale stems from the length scale, the size of the hydrogen atom must be on the order of an Å. The same holds true for the size of all heavier atoms, for reasons we already mentioned (Section 1.3.5).

Based on this, we also need to introduce a suitable terminology to reflect the fundamental role in quantum physics of the *allowed*—or *quantized—energies* of a quantum system. Thus, we now speak of *energy levels* of the system and depict them in the so-called *energy-level diagram*, as in Figure 1.8.

1.3.9 Correct and Incorrect Pictures for the Motion of Electrons in Atoms: Revisiting the Case of Hydrogen

After our discussion so far, how can we picture the hydrogen atom, at least in its ground state? Which description would be consistent with the wave nature of the electron and at the same time devoid of concepts with no experimental meaning, like the electronic orbit? The answer is plain to see. The picture we seek should be the analog of a classical standing wave with a similar geometry; for example, a classical sound wave in the interior of a hollow sphere filled with air. The simplest form of such a wave—corresponding to the "fundamental"

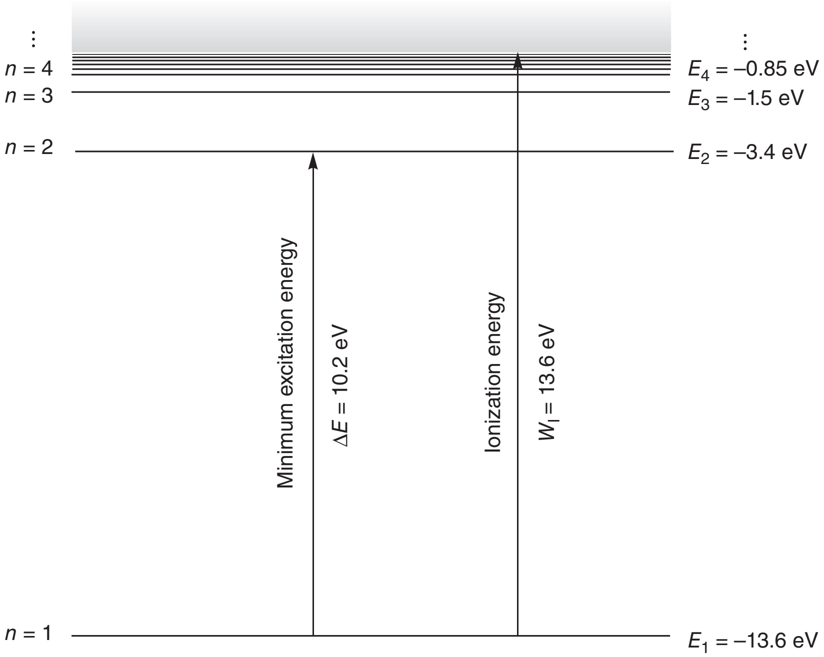


Figure 1.8 The energy-level diagram for the hydrogen atom and the two basic quantities associated with it. The ionization energy W_I is the minimum energy needed to remove the electron from the atom. The minimum excitation energy ΔE is the energy required to affect the smallest possible change to the atom in its ground state.

frequency—is a spherically symmetric compression (high pressure) that periodically becomes a rarefaction (low pressure), while retaining its spherical shape. Such a wave has a time dependence of the form

$$p(r,t) = p(r)\cos\omega t, \tag{1.36}$$

where the pressure p is measured with respect to atmospheric pressure—that is, a positive sign refers to higher-than-atmospheric and a negative sign to lower-than-atmospheric pressure.

Note that, just like in the case of a string, the following is also true for the normal oscillation modes of two- or three-dimensional objects: The shape of the fundamental oscillation has no *nodes* (i.e., no *nodal lines* for two-dimensional objects and no *nodal surfaces* for three-dimensional ones), while it also has the full symmetry of the problem. In a string, the fundamental oscillation is symmetric with respect to its midpoint; in a two-dimensional object with circular geometry, such as a drum, it is rotationally symmetric with respect to its center; and likewise for three-dimensional objects with spherical symmetry. It follows that the "fundamental" standing electron wave around the nucleus is also spherically symmetric with no nodal surfaces.

The answer, therefore, to the question "how are we to picture the hydrogen atom in its ground state," is depicted in Figure 1.9: a spherically symmetric cloud of probability density that engulfs the nucleus. Such a cloud roughly

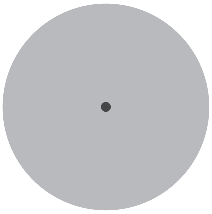


Figure 1.9 The correct quantum picture for the ground state of the hydrogen atom. The wave nature of the electron is incompatible with motion along some classical orbit. Instead, we are forced to think of the electron (in its ground state) as a spherically symmetric *probability cloud* about the nucleus.

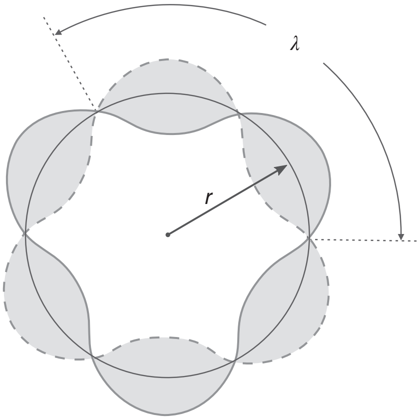


Figure 1.10 A false picture that should be discarded. Here the electron supposedly forms something like sinusoidal standing matter waves along a circle of radius *r*. But this picture is a flawed projection to three-dimensional space of the classical picture for a wave on a string. Three-dimensional waves—quantum or classical—typically fill the space and surely do not look like standing sound waves in a circular tube.

represents—according to the probabilistic interpretation of matter waves we mentioned earlier—the region of space where it is highly likely to find the electron. (In reality, the wave extends outside the shaded region, but with exponentially diminishing amplitude.)

But unlike pressure fluctuations in a classical gas (that can vanish periodically as compressions become rarefactions), a probability "compression," as in Figure 1.9, cannot periodically disappear, since the electron it describes would then also disappear! It follows that quantum waves—precisely because of their interpretation as probability waves—cannot evolve temporally as in (1.36), but must instead have a different time dependence, which we shall unravel later.[6] For the time being, let us retain the notion that quantum waves are similar to classical waves in their spatial form, but distinctly different with respect to their physical interpretation and time evolution.

The preceding discussion focused on a rudimentary description of the most basic quantum system, namely, the electron in the ground state of the hydrogen atom. But it also serves another purpose in helping us eliminate the false picture of Figure 1.10 for the hydrogen atom.

---

[6] Instead of $\cos \omega t$ in (1.36), the time dependence has the complex form $\exp(-i\omega t)$, where $\omega = E/\hbar$. But since $|e^{-i\omega t}| = 1$, such a time dependence implies that the probability distribution of the electron around the nucleus (in a state of a given energy and frequency) remains unchanged in time. For the ground state in particular, this means that the physical and chemical properties of the atom remain invariant in time, as one should expect. (By the way, this is the reason atoms in their ground state do not radiate.) The complex form of the time evolution is thus a crucial difference between quantum and classical waves, without which the physical interpretation of the former would be impossible.

According to this false picture—used extensively in many textbooks, due to its success in explaining Bohr's quantization condition (see subsequent text)—the standing electronic waves of the hydrogen atom are formed along a circle whose radius r satisfies the relation

$$2\pi r = n\lambda. \tag{1.37}$$

In other words, electronic waves are formed if an integer number of wavelengths fits on the said circle. But since $\lambda = h/p$, expression (1.37) can be written as $rp = nh/2\pi$, or, equivalently,

$$\ell = n\hbar, \qquad (\ell = rp = mvr), \tag{1.38}$$

which is the celebrated *Bohr's quantization condition* (Bohr, 1913): The electron in the hydrogen atom can only move along specific quantized circular orbits for which its angular momentum, $\ell = mvr$, is an integer multiple of Planck's constant, $\hbar$. Equation (1.38) together with Newton's law for a circular orbit,

$$m\frac{v^2}{r} = \frac{e^2}{r^2} \qquad \text{(cgs)}, \tag{1.39}$$

form a system of two equations with two unknowns, v and r. Its solution yields

$$r_n = n^2\frac{\hbar^2}{me^2} = n^2 a_0 \qquad \text{(allowed radii)} \tag{1.40}$$

$$v_n = \frac{e^2}{\hbar}\frac{1}{n} \qquad \text{(allowed speeds)} \tag{1.41}$$

$$E_n = -\frac{me^4}{2\hbar^2}\frac{1}{n^2} \qquad \text{(allowed energies)}, \tag{1.42}$$

where $a_0 = \hbar^2/me^2 = 0.529\,\text{Å} \approx 0.5\,\text{Å}$ is the so-called *Bohr radius,* which correctly predicts the size of the atom in its ground state ($n = 1$). Likewise, the quantity $W_\text{I} = me^4/2\hbar^2$ in the energy formula gives the ionization energy of the atom and its numerical value (13.6 eV) agrees with our earlier finding using experimental data.

But despite the empirical success of Bohr's theory—and its theoretical justification based on de Broglie waves—we should not give in to the tempting thought that Bohr's theory describes something real. There are no quantized orbits—since no orbits exist at all—and no standing de Broglie waves of the type shown in Figure 1.10. As we have seen, this picture is *flawed*; therefore, the sooner we put it aside, the better.[7]

---

[7] It is curious that this flawed picture appears to go back all the way to the time of Schrödinger and de Broglie! (Which may partly explain its endurance, despite its falseness.) Indeed, as recounted by Felix Bloch (Section 2.2), Schrödinger himself had used this picture to obtain Bohr's quantization rules, an approach for which he was chided by Debye who characterized this way of thinking as "childish," thus prodding Schrödinger to delve deeper and come up with his eponymous wave equation. So, to a small extent, we may owe the discovery of the Schrödinger equation to this flawed picture of waves fitted along a stationary orbit. Actually, waves of this form can also appear in modern quantum mechanics, but only in the, so-called, classical (or semiclassical) limit of the theory.

Let us now see how we can use the correct three-dimensional picture of the spherical probability cloud for the hydrogen atom (Figure 1.9), in order to view in the right context the approximate condition $\lambda \sim 2L$ we frequently invoked to explain the basic features of the atomic world. For a spherical probability cloud we can say the following. Although the concept of the wavelength has meaning only for sinusoidal waves, we can introduce a sort of wavelength for three-dimensional standing waves (which are anything but sinusoidal), by defining the distance between adjacent nodes as half-wavelength, or the distance between a peak and the nearest node as quarter-wavelength. Therefore, for a spherical probability cloud with a peak at $r = 0$ and the node at $r = a$, we obtain the approximate expression $\lambda/4 \approx a \Rightarrow \lambda = 4a$. We could arrive at the same result by saying that a diameter of the sphere is half a wavelength ($\lambda/2 = 2a \Rightarrow \lambda \approx 4a$) because the wave vanishes at the endpoints of the diameter.

1.3.10 The Fine Structure Constant and Numerical Calculations in Bohr's Theory

Although we emphasized that Bohr's theory of quantized orbits does not provide a correct picture of the atom—since the wave nature of electrons excludes the existence of orbits—the theory is nevertheless useful for quick calculations that give us a first quantitative description of the essential features of the hydrogen atom. These calculations are further simplified if we use the so-called *fine structure constant*, which is defined, in cgs units, as

$$\alpha = \frac{e^2}{\hbar c} \approx \frac{1}{137} \qquad \text{(fine structure constant)}, \tag{1.43}$$

and is a dimensionless quantity, as already mentioned in Section 1.2.3. Using (1.43) we can rewrite formula (1.41) as

$$\upsilon_n = \frac{e^2}{\hbar c}\frac{c}{n} = \frac{\alpha c}{n} \qquad (n = 1, 2, \ldots), \tag{1.44}$$

which tells us that the speed of the electron in the first Bohr orbit ($n = 1$) is approximately 137 times smaller than the speed of light. Even though this small value justifies, to first order, the nonrelativistic treatment of the problem, it also shows that relativistic effects are not negligible to second order. Such effects are manifested particularly as small shifts in the energy levels of the atom—and the corresponding spectral lines—and thus produce the so-called *fine structure* of the spectrum. The name of the quantity (1.43) refers exactly to this fine structure.

The constant α is also quite useful for quick numerical calculations, since it allows us to eliminate from the pertinent expressions the square of the electric charge e^2 by setting

$$e^2 = \alpha \hbar c, \tag{1.45}$$

so that e^2 is given in terms of more familiar physical constants. Actually, the substitution (1.45) in various expressions allows us to perform the calculation in a quick and elegant manner. In the case of ionization energy, for example, we have

$$W_{\mathrm{I}} = \frac{me^4}{2\hbar^2} = \frac{1}{2}\frac{m(\alpha\hbar c)^2}{\hbar^2} = \frac{1}{2}\alpha^2 mc^2 = \frac{1}{2}\left(\frac{1}{137}\right)^2 0.5\,\mathrm{MeV} = 13.6\,\mathrm{eV}.$$

As you see, we are able to perform the calculation promptly—and in the appropriate energy unit of eV—by simply using the rest energy of the electron (a value worth remembering, since it is customary to refer to particle masses not in units of g or kg, but in terms of their equivalent energy). We also note that the factor α^2 before mc^2 alerts us about the order of magnitude of atomic energies: They are some five orders of magnitude smaller than an MeV, that is, on the order of 10 eV.

In practice, it also pays to remember the numerical values of some equivalent combinations of parameters of the hydrogen atom (e.g., m, e, and $\hbar$) with the dimension of energy. Three such equivalent combinations that arise often are $me^4/\hbar^2$, $\hbar^2/ma_0^2$, and e^2/a_0, where a_0 $(= \hbar^2/me^2)$ is the Bohr radius. They are equal to each other and have twice the value of the ionization energy of the atom, namely, 27.2 eV. As we shall see later—when we introduce the so-called *atomic units*—this value is the natural unit of energy in the atomic world; it is known as *one Hartree* or the *atomic unit of energy*.

In the following example, we describe how to calculate in a quick and transparent way another important property of Bohr's theory: the intensity of the electric field acting on the electron in its ground state.

Example 1.4 Calculate, in SI units, the electric potential and the electric field intensity at a distance of one Bohr radius from the nucleus of the hydrogen atom.

Solution: The electric potential is obviously $V = -27.2\,\text{V}$, since the corresponding potential energy of the electron is $V = -e^2/a_0$ and the absolute value of this quantity is 27.2 eV, that is, 27.2 e · V. For the intensity $\mathcal{E}$ of the electric field we should note that the pertinent unit in the SI system is the volt per meter (V/m), given that the product $\mathcal{E} \times$ distance yields the potential difference between two points. In the present case we have

$$\mathcal{E} = \frac{e}{a_0^2} \equiv \frac{e^2}{a_0}\frac{1}{e}\frac{1}{a_0} = 27.2\,\text{eV} \cdot \frac{1}{e} \cdot \frac{1}{0.5 \times 10^{-10}\,\text{m}} = 5.44 \times 10^{11}\,\text{V/m}.$$

Here we simply rewrote the initial expression e/a_0^2 to form the energy combination $e^2/a_0 = 27.2\,\text{eV}$ and then divided by the electric charge e to isolate the volt. Finally, we divided by the Bohr radius $a_0 = 0.5 \times 10^{-10}$ m to obtain V/m, which is the unit of the electric field in the SI system.

It should be evident by now how cumbersome the conventional systems of units—SI or cgs—are for calculations in the atomic world. If there is any shred of doubt remaining, we encourage the readers to attempt the preceding calculation in any one of those systems. Good luck!

In the following example we expand Bohr's theory to the so-called *hydrogen-like atoms* (or *ions*).

Example 1.5 Calculate (in eV) the ionization energies of the first three *hydrogen-like atoms*—He^+, Li^{++}, and Be^{+++}—together with their corresponding radii.

Solution: Let us first clarify that the so-called hydrogen-like atoms are simply ions of heavier elements that have been stripped of all but *one* of their electrons. A hydrogen-like atom is thus identical to hydrogen, except that it has Z protons in its nucleus. Since the nuclear charge is then Ze, the Coulomb force on the electron is equal to $F = (Ze) \cdot e/r^2 \equiv Ze^2/r^2$, instead of e^2/r^2 for hydrogen. Clearly, applying Bohr's theory to any given hydrogen-like atom with atomic number Z yields the "same" results as for hydrogen, albeit with Ze^2 instead of e^2. For example, for the ionization energy $W_{\mathrm{I}} = me^4/2\hbar^2$ we obtain

$$W_{\mathrm{I}}(Z) = W_{\mathrm{I}}(\mathrm{H})|_{e^2 \to Ze^2} = \frac{mZ^2e^4}{2\hbar^2} = Z^2 W_{\mathrm{I}}(\mathrm{H}) = Z^2 \cdot 13.6 \text{ eV},$$

while for the corresponding radii we have

$$a_0(Z) = a_0(\mathrm{H})|_{e^2 \to Ze^2} = \frac{1}{Z}\frac{\hbar^2}{me^2} = \frac{1}{Z} \cdot 0.5\,\text{Å}.$$

In particular, for the ionization energies of He^+ (singly ionized helium), Li^{++} (doubly ionized lithium), and Be^{+++} (triply ionized beryllium), we find

$$W_{\mathrm{I}}(\mathrm{He}^+) = Z^2 \cdot 13.6\,\text{eV}|_{Z=2} = 54.4\,\text{eV},$$
$$W_{\mathrm{I}}(\mathrm{Li}^{++}) = 122.4\,\text{eV}, \quad W_{\mathrm{I}}(\mathrm{Be}^{+++}) = 217.6\,\text{eV},$$

in excellent agreement with experimental data. Moreover, the complete set of allowed energies of a hydrogen-like atom is given by the formula

$$E_n(Z) = E_n(\mathrm{H})|_{e^2 \to Ze^2} = Z^2 E_n(\mathrm{H}) = -\frac{Z^2 \cdot 13.6}{n^2} \text{ eV}.$$

For instance, in the case of He^+ we obtain

$$E_n(\mathrm{He}^+) = -\frac{54.4}{n^2} \text{ eV}.$$

This prediction is in spectacular agreement with the "dark spectral lines" in the sun's absorption spectrum, which are therefore attributed to the existence of He ions in the solar atmosphere. Actually, the successful explanation of these lines was one of the early triumphs of Bohr's theory and played a decisive role in its adoption by the scientific community. Nevertheless, a few years later, the wave–particle duality of matter was discovered, leading to the realization that Bohr's theory, although a successful calculational model for hydrogen-like atoms, was based on notions (such as quantized orbits) that have no place in a modern quantum mechanical context.

1.3.11 Numerical Calculations with Matter Waves: Practical Formulas and Physical Applications

To facilitate calculations with matter waves, we first need to rewrite formula $\lambda = h/p$ so that it gives us the wavelength of a particle directly in Å, once we have its energy in eV. Indeed, what we typically know is not the speed or momentum of, say, an electron, but its energy in eV, since we normally deal with a beam of such particles that have been accelerated by some potential difference expressed in volts. So, for this purpose, we rewrite formula $\lambda = h/p$ as

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE}} = \frac{hc}{\sqrt{2mc^2 \cdot E}} = \frac{12\,400\,\text{eV Å}}{\sqrt{2 \cdot 0.5 \times 10^6\,\text{eV} \cdot E(\text{eV})}},$$

where we invoked the nonrelativistic energy–momentum formula $E = p^2/2m \Rightarrow p = \sqrt{2mE}$ and multiplied the numerator and denominator by the speed of light c. We thus obtained mc^2 in the denominator and hc in the numerator. The latter is a constant whose value, in eV Å, is already known to us. (It appears in the relation $\epsilon = hf = hc/\lambda$ for photons, where we found that $\epsilon(\text{eV}) = 12\,400/\lambda(\text{Å})$ and therefore $hc = 12\,400\,\text{eV Å}$.) We thus obtain the following practical formula:

$$\lambda_e(\text{Å}) = \frac{12.4}{\sqrt{E(\text{eV})}} \approx \frac{12}{\sqrt{E(\text{eV})}}, \tag{1.46}$$

while for the proton and the neutron (i.e., the nucleons) we have

$$\lambda_{\text{N}}(\text{Å}) = \lambda_e(\text{Å})\sqrt{\frac{m_e}{m_p}} = \frac{\lambda_e(\text{Å})}{\sqrt{1836}} \approx \frac{\lambda_e(\text{Å})}{43}. \tag{1.47}$$

Here we took into account that $m_p \approx m_n \approx 1836\,m_e$, and that the dependence of the wavelength on the mass is $1/\sqrt{m}$. According to (1.47), the de Broglie wavelength of a nucleon is about 43 times smaller than the corresponding wavelength of an electron of the same energy. Taking (1.46) into account, the corresponding formula for nucleons becomes

$$\lambda_{\text{N}}(\text{Å}) = \frac{0.289}{\sqrt{E(\text{eV})}} \approx \frac{0.3}{\sqrt{E(\text{eV})}}. \tag{1.48}$$

A simple order-of-magnitude test of formula (1.46) is the following: For electrons in the ground state of the hydrogen atom, whose kinetic energy is 13.6 eV, the wavelength should be roughly twice the atomic diameter—that is, on the order of a few angstroms. Indeed, (1.46) confirms this expectation.

An interesting practical conclusion emerges from the comparison of formula (1.46) and the corresponding one for photons; namely, $\lambda_\gamma(\text{Å}) = 12\,400/\epsilon(\text{eV})$: To attain a wavelength on the order of a few angstroms, an electron needs an energy of a few eV, while a photon needs a few thousand eV, respectively. A direct implication of this fact is the concept of the *electron microscope*. As is known from optics, the resolution of a microscope is limited by the wavelength of the light being used. For a given λ, we cannot see any details of the object if their size d is less than or equal to λ. To obtain a resolution of a few angstroms with an "optical" microscope, we would need to employ photons of a few thousand eV—that is, x-rays—which are hard to manipulate, for example, focus.

However, if our microscope uses electrons instead of photons—that is, matter waves instead of EM ones—then a resolution of a few angstroms can be achieved with electrons of very low energy (a few eV), which can readily be used in many applications. In fact, if our desired resolution is 100 Å—which is good enough for most biological uses of the electron microscope—then the required energy decreases to a few hundredths of an eV, a nondestructive energy for the sample examined. (It does not cause ionizations or dissociations of chemical bonds, etc.) At the same time, the manipulation of an electron beam is feasible using an appropriate system of electric and magnetic lenses.

Another important application of matter waves pertains to the so-called *thermal neutrons*. These are produced in large quantities in nuclear reactors when initially fast-moving neutrons impinge on a material containing *light nuclei*[8] (e.g., graphite). In this way, neutrons slow down through collisions with the graphite nuclei and eventually reach thermal equilibrium. Their kinetic energies are then approximately equal to $E(\text{eV}) \approx (kT)_{T\approx 300\,\text{K}} \approx 1/40\,\text{eV}$, so formula (1.48) yields

$$\lambda_n(\text{Å}) \approx \frac{0.3}{\sqrt{E(\text{eV})}} \approx 0.3 \cdot \sqrt{40} \approx 2\,\text{Å}.$$

This result means that thermal neutrons are the prototypical *matter wave* for crystallographic studies, because, first, they interact with lattice nuclei (and are therefore sensitive to their positions) and, second, internuclear distances are on the same order of magnitude as the wavelength of these neutrons (so the required resolution is available).

Going to the other extreme—the macroscopic world—let us now calculate the wavelength of a dust particle, which has a mass of 1 mg and is moving at a speed of 1 cm/s. In this case, we have (in cgs units)

$$\lambda = \frac{h}{mv} \approx \frac{6 \times 10^{-27}}{10^{-3} \cdot 1} = 6 \times 10^{-24}\,\text{cm}.$$

To observe the wave nature of a “particle” that has such a tiny wavelength, we need interference or diffraction experiments with slits or obstacles of a size similar to that of the wavelength. But neither of these exists in nature, since even the smallest “object” we know of—the atomic nucleus—has dimensions of 10^{-13} cm.

We therefore conclude that, even though the principle of wave–particle duality of matter is in theory also applicable to the motion of macroscopic bodies, it has no measurable consequences for them and can thus be ignored in practice.

1.3.12 A Direct Confirmation of the Existence of Matter Waves: The Davisson–Germer Experiment

As we noted at the beginning of this section, the direct confirmation of the existence of matter waves requires an interference experiment. The simplest such experiment was conducted for the first time—somewhat inadvertently—in 1927 by Clinton Davisson and Lester Germer. When they bombarded a nickel crystal with a monoenergetic beam of electrons with $E = 54\,\text{eV}$, they were surprised to discover that apart from the normal reflection, the beam had a preferential angle of oblique reflection equal to 51° with respect to the direction of incidence. The crucial thing to note here is that—due to their low energy—electrons do not penetrate the crystal, and reflection, therefore, takes place only at the surface. Figure 1.11 depicts the main idea of the experiment.

If we adopt the de Broglie hypothesis, then the theoretical analysis of the experiment is simple. The incident electron wave is reflected at the locations of the Ni atoms in the form of secondary spherical wavelets with the same

---

[8] Because only then there is a sizable energy transfer from the neutrons to the nuclei.

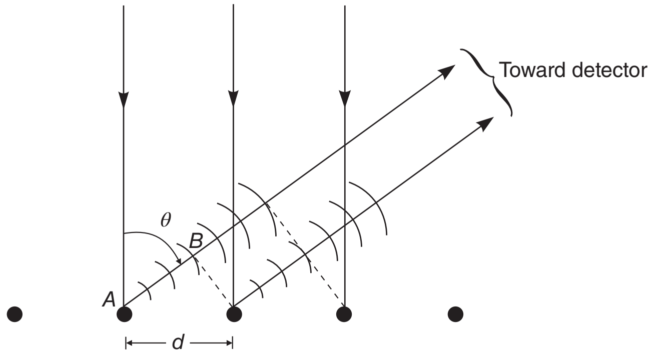


Figure 1.11 Theoretical analysis of the Davisson–Germer experiment. The electrons are scattered preferentially toward those directions θ that satisfy the condition of constructive interference $d \sin\theta = n\lambda$ ($n = 0, 1, 2, \ldots$).

wavelength as the initial wave. These wavelets interfere constructively only in those directions θ for which the path difference $AB (= d \sin\theta)$ between two adjacent reflected "rays" is an integer multiple of the wavelength λ of the electrons. In other words, constructive interference occurs when

$$d \sin\theta = n\lambda. \tag{1.49}$$

In the case at hand, we have $d = 2.15\,\text{Å}$ (this was already known from earlier diffraction measurements of crystalline nickel using x-rays) and therefore, according to formula (1.46) for $E = 54\,\text{eV}$,

$$\lambda(\text{Å}) = \frac{12.4}{\sqrt{54}} = 1.68\,\text{Å}.$$

Thus, condition (1.49) yields

$$\sin\theta_n = n\frac{\lambda}{d} = n\frac{1.68}{2.15} = 0.78n,$$

whence we see that, apart from the trivial case of normal reflection ($n = 0 \Rightarrow \theta_0 = 0$), there is only one more scattering direction at an angle

$$\sin\theta_1 = 0.78 \Rightarrow \theta_1 = 51^\circ,$$

just like the experiment revealed! Thus, from 1927 onward de Broglie waves were no longer a theoretical conjecture, but an irrefutable experimental fact.

1.3.13 The Double-Slit Experiment: Collapse of the Wavefunction Upon Measurement

The evidence we presented so far in support of the principle of wave–particle duality of matter is so compelling that one may wonder why there would be a need at all for yet another pertinent experiment. Let us therefore stress right away that the double-slit experiment—originally a thought experiment that was subsequently conducted numerous times—is not included here as just another confirmation of the dual nature of particles. It serves, rather, as an ideal "tool" for investigating the central concept of quantum theory, namely, the probabilistic interpretation of matter waves. Such an interpretation is compulsory for reasons we already explained (Section 1.3.6). It provides the only conceivable way to combine, in the same physical entity, the mutually exclusive properties of being *localized and indivisible* (akin to particles) on the one hand, and *extended and divisible* (akin to waves) on the other. By interpreting the wave associated with a particle as a probability wave, we no longer deny the corpuscular nature of the particle. The wave here simply describes the probability of finding the particle here or there, but never here *and* there at the same time. This implies that quantum particles (e.g., electrons) are always detected as integral and indivisible entities. They leave, for example, point-like traces on a photographic plate.

However, the "abstract" nature of quantum waves—we could also call them *information waves* or even *waves of knowledge* (of probabilistic nature) pertinent to a particle's state—has as an inescapable consequence the so-called *collapse of the wavefunction upon measurement*. This effect underlies all major paradoxes of quantum mechanics, such as those arising from the double-slit experiment we shall discuss shortly. The collapse of the wavefunction highlights the central role of the measurement process in the quantum world. What this term means is somewhat self-evident: If a measurement yields some information about the particle, then its wavefunction immediately after the measurement must reflect what we just measured, and must "incorporate" the information obtained from the measurement. That is, if we were to repeat the measurement a second time—on the same particle that was just observed—we should always confirm the first measurement. For example, if we measure the position of a particle—say, in one dimension, for simplicity—and locate the particle in the vicinity of a point x_0, then its wavefunction after the measurement can only be highly localized about the point x_0. The said wavefunction thus represents a particle whose location is now known to us with high precision—or at least with as much precision as we were able to measure it. The result of the measuring process is thus an *abrupt*—in fact, *instantaneous*—"shrinking" of the initial wavefunction to its new form that is dictated by the result of the measurement. We depict all this in Figure 1.12.

The "collapse" of the wavefunction so that it "adapts" to experimental data is thus an inescapable logical consequence of the probabilistic interpretation. Without this collapse the probabilistic interpretation would make no sense.

We are now ready to discuss the double-slit experiment in the abovementioned context.

Two alternative pictures of the experiment are shown in Figure 1.13. In the first picture, a beam of particles with prescribed momentum—and hence, wavelength—impinges on a plate with a double slit. The beam is depicted as a plane wave that emerges on the other side of the plate as two circular wavelets centered on the two slits. An image of interference fringes appears on the screen—a sort of photographic plate that records the particles arriving at its various regions—depending on whether the two circular wavelets emanating from the slits arrive in the specific region in phase or out of phase. In the second picture, the impinging beam is shown as a "stream of particles." The particles fall on the plate, go through one or the other slit, and form interference fringes

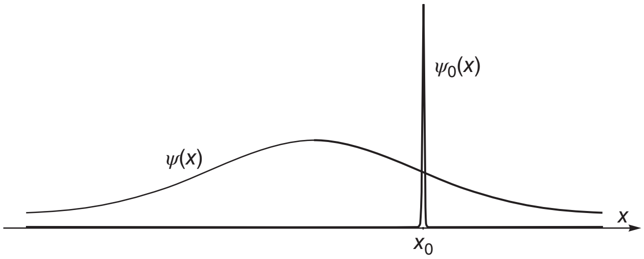


Figure 1.12 The “collapse” of the wavefunction in a position measurement: When a measurement detects a particle at position x_0, its wavefunction collapses immediately to a highly localized form around x_0 and instantaneously vanishes elsewhere. It is as if the measurement “sucks” the wavefunction, only to concentrate it suddenly at the point where the particle was located. Clearly, such an instantaneous collapse has a nonlocal character; that is, it seems to imply some sort of *action at a distance*. But since the wavefunction is a mathematical entity—not a physical wave with energy and momentum distributed in space—this instantaneous collapse does not imply a corresponding instantaneous transfer of energy or momentum and therefore it does not violate the theory of relativity. A measurement simply “removes” all possibilities to locate the particle anywhere else than the position it was found to be.

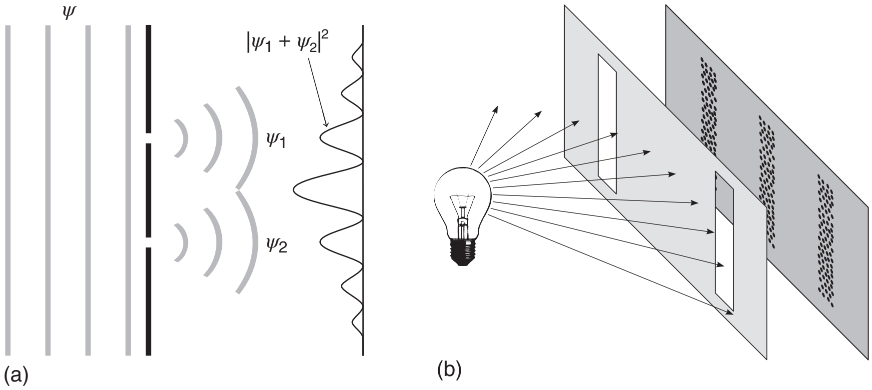


Figure 1.13 Two alternative pictures for the double-slit experiment: (a) The *wave picture*. (b) The *particle picture*. Both pictures are legitimate. But only the wave picture—with the understanding of the wave as a probability wave—provides a qualitatively and quantitatively correct understanding of this experiment. The particle picture is just to remind us that there are only particles “behind” the wave that describes how they move in space.

as before, by appearing in various regions in the screen in smaller or greater numbers. After all, it is the same experiment.

But in contrast to the wave picture—which predicts naturally the interference fringes of the actual experiment—the particle picture can never lead us to the correct description of the phenomenon. The reason is that the particle picture implies from the outset the false notion that the particles move in classical orbits. On such a flawed basis, no valid predictions can be drawn, especially about interference fringes. The conclusion is completely general: The only basis for the proper description of quantum phenomena is always the wave picture, with the additional clarification that the corresponding waves are to be interpreted as probability amplitude waves,[9] not as classical waves. We thus arrive at the key point for the double-slit experiment. Let us view it from the perspective of a person who is skeptical of the statistical interpretation, and whose line of thinking is the following: "There is no doubt that the incident beam behaves like a wave, since it produces the expected interference fringes on the screen. But how do we know that this is a probability amplitude wave and not a truly classical wave? Can we experimentally distinguish these two possibilities, since they both lead to the same interference fringes?" This is the real conceptual question that needs to be clarified by the double-slit experiment.

Let us begin then. A probability wave—if this is what is going on here—does not represent a measurable physical disturbance. It takes physical meaning only through the particle it describes: An experiment can only detect the particle and nothing else. The experimental question pertinent to the particle is clearly the following: Which slit did it go through? Clearly, the particle can pass through *either* one slit *or* the other (but not through both) because in the context of wave–particle duality (and its probabilistic interpretation) particles are integral, indivisible entities, and are always detected as such.

Therefore, if we were to place two detectors near the exit of each slit—so that they register an event when a particle passes through—then we expect the following experimental outcome (provided that the incident beam is so dilute that only one particle arrives at the plate at any one time): (i) Only one of the two detectors will register an event every time. This means that it is only through one slit that "something" goes through at that instant—obviously, this is the particle—while nothing at all goes through the other slit. (ii) If we repeat this experiment many times and measure how many times a particle passes through one slit or the other, the two numbers will tend to be equal. The reason for this is that the incoming plane wave has the same amplitude at the entry point of each slit and therefore the probabilities of locating the particle at the entry of one or the other slit are equal.

But if the waves in our experiments were classical waves of some kind—that is, if they represented some measurable physical disturbance—then the two detectors would continuously register an event, since in this case, a physical wave (not a probability wave) would be transmitted continuously through both slits. It is clear, therefore, that if we confirm the prediction of the probabilistic interpretation—that something passes through one slit and nothing passes through the other—then the classical picture of a continuous passing of the wave through both slits fails and is thus rejected. The experiment actually confirms the quantum mechanical prediction. It is only through one slit at a time that something passes, while the transmission numbers are indeed equalized eventually.

But the most fundamental difference between the two types of waves—the probability waves and the classical waves—is what we are about to describe now. If the waves in our experiment are indeed probability waves, then the placement of detectors immediately behind the slits would have a dramatic consequence:

---

[9] The term *amplitude* is necessary because the wave nature is represented by the wavefunction ψ, that is, by the *wave amplitude.* But for brevity we will often use the term *probability wave* instead of *probability amplitude wave.*

Interference fringes disappear! By contrast, if the waves are classical then the presence of the detectors should not affect the interference fringes!

For classical waves, the abovementioned statement is self-evident. In the classical world the act of observation can always be made so as not to affect the observed phenomenon in any significant way.

It is therefore impossible to explain in classical physics the disappearance of the interference fringes due to the presence of the two detectors. And yet this disappearance is an experimental fact. All pertinent experiments confirm it beyond any doubt. As soon as we activate the detectors to alert us as to which slit the particle passed through, the interference fringes vanish!

A straightforward explanation of the disappearance of interference fringes is provided by the collapse of the wavefunction due to measurement. We remind the readers that in the context of the statistical interpretation, the wave describing the particle is not a physical wave but an "information wave," and, as such, it should always respond instantaneously to the "new knowledge" about the particle that is obtained via measurement. In other words, the probability (or information) wave ought to express what we know at every instance.

Here is the direct implication of this thinking on our experiment (Figure 1.14). As soon as the detector observes the particle passing through slit #1, the probability to detect it simultaneously at slit #2 vanishes and the corresponding wavelet (in the vicinity of slit #2) disappears at once. The probability wave immediately after the measurement—where transmission through slit #1 was recorded—contains only the circular wavelets centered on that slit. A similar statement can be made when the other detector records transmission through slit #2: We would have circular wavelets around slit #2, and nothing around slit #1. If we were dealing with classical waves, the only way to obtain such an outcome would be to shut one or the other slit at a time. As a result, there would be no interference fringes on the screen (since the two slits would never be open at the same time), but a mere merging of the two diffraction patterns[10] around the projections of the slits onto the screen. Indeed, this is what we obtain in our experiment when the detectors near the slits are in operation, informing us from which slit the particle passed through every time. But as soon as we switch the detectors off, interference fringes reappear in all their glory!

Let us also note that the quantum nature of the incident wave is revealed clearly from the way the interference fringes form in the course of the experiment. The fringes appear gradually as the "spots"—that is, the traces of the particles impinging on the screen—accumulate. And while all such particles are described by the same quantum wave—and are thus in no way different from one another—we observe that each particle "lands" at a different spot on the screen. Each such event is fundamentally unpredictable and nobody can say why a particle falls on a particular spot. Yet all these events together must form the probability distribution described by the intensity of the quantum wave in each region of the

---

[10] In a typical diffraction experiment, a wave beam that passes through a tiny slit emerges in wide angles if the size of the slit is comparable to or narrower than the wavelength of the beam. Because of the beam's angular opening, the trace of the diffracted beam on a screen is not only centered at the projection of the slit but also extends away from it with decreasing intensity.

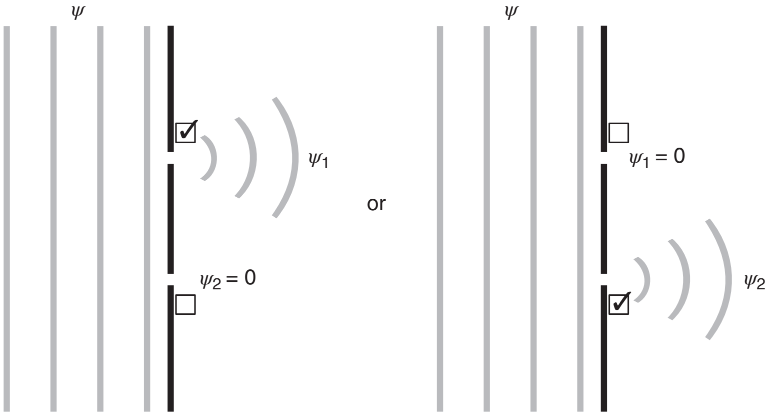


Figure 1.14 Quantum mechanical explanation of the disappearance of interference fringes. Because of the measurement, the wavefunction collapses into the wave that passes through the slit where detection occurs. Simultaneous "emission" of probability waves from both slits is no longer possible, and interference fringes disappear.

screen. This intensity is high in areas where the two circular wavelets interfere constructively and low in areas where the two wavelets interfere destructively.

We conclude the section by examining how we got here. We initially stated our twofold goal: to empirically check the principle of wave–particle duality—this fundamental principle of quantum theory—and to familiarize our readers with it. Our aim was to make them view the principle not as a whim of nature (something we have to accept simply because it is empirically correct), but as the only natural explanation of the most challenging mysteries of the atomic world: the inexplicable, from a classical viewpoint, stability of its structures (say, atoms and molecules) and the uniqueness of their form: The fact that no matter how many times we "break up" an atom or a molecule and allow it to form anew, it always emerges in identical form. In other words, the microscopic constituents of nature—atoms, molecules, nuclei, and so on—have no history. Their form is predetermined—like that of normal oscillation modes in classical standing waves—and does not depend on how and when they were created. In the words of Niels Bohr:

> I had best begin by telling you a little about the history of this theory. My starting point was not at all the idea that an atom is a small-scale planetary system and as such governed by the laws of astronomy. I never took things as literally as that. My starting point was rather the stability of matter, a pure miracle when considered from the standpoint of classical physics. By 'stability' I mean that the same substances always have the same properties, that the same crystals recur, the same chemical compounds, etc. In other words, even after a host of changes due to external influences, an iron atom will always remain an iron atom, with exactly the same properties as before. This cannot be explained by the principles

> of classical mechanics, certainly not if the atom resembles a planetary system. Nature clearly has a tendency to produce certain forms—I use the word 'forms' in the most general sense—and to recreate these forms even when they are disturbed or destroyed.
>
> [Werner Heisenberg, *Physics and Beyond: Encounters and Conversations*, Translated by Arnold J. Pomerans, Harper & Row (New York, 1971).]

We encouraged the readers earlier, and emphatically continue to do so now, to return to the "mystery of the atomic stability" every time they feel intimidated by the paradoxical features of quantum theory. When they realize time and again the scandalous failure of classical physics to explain this mystery, they can retrace the chain of thought we put forward in Section 1.3.2 and summarize again here as follows:

$$\begin{aligned}&\text{Stability} \rightarrow \text{Quantization} \rightarrow \text{Wavelike behavior} \rightarrow \text{Wave}\\ &\quad = \text{Probability wave} \rightarrow \text{Collapse of the quantum wave upon measurement.}\end{aligned}$$

We leave it to the readers to traverse this chain of thought—as a kind of *conceptual exercise*—by bringing forward the arguments we developed earlier. There is not much room for alternatives. Based on our knowledge today, it is clear that the fundamental principles of quantum mechanics arise readily as the only natural explanation of the mystery of atomic stability and the uniqueness of atomic structures.

Actually, it is classical physics—not quantum mechanics—that ought to shock us when we try to explain the phenomena of the atomic world.

Problems

1.5 To familiarize yourselves with the "practical formulas" of wave–particle duality, calculate the following quantities:

(a) The de Broglie wavelength of an alpha particle accelerated by a potential difference of 50 V.

(b) The energy of the incident electron beam in a Davisson–Germer experiment, where the crystal has $d = 2.48$Å and the angle of maximum oblique reflection is $\theta = 30°$.

(c) The de Broglie wavelength of protons in a CERN experiment (CERN is the European Organization for Nuclear Research near Geneva, Switzerland), where their energy is on the order of 10 TeV $= 10^{13}$ eV. Show first that for ultrarelativistic particles, whose rest energy is practically negligible compared to their kinetic energy, the formula for the wavelength is the same as for photons. Why is this to be expected?

1.6 (Particle in a tubule.) For some linear organic molecules the valence electrons can practically move freely along the molecule's main axis, without being able to leave the molecule. We can therefore approximate their motion with that of a free-moving particle—of mass m—inside a tubule of length L (equal to the molecule's length) from which it cannot exit. The particle is permanently trapped there. Apply the condition for the formation of standing waves to calculate the allowed energies of the particle trapped in the tubule.

1.7 What do you expect will happen to the Bohr radius ($a_0 = \hbar^2/me^2$) and the ground-state energy of the hydrogen atom ($E_1 = -me^4/2\hbar^2$) in the following limits:

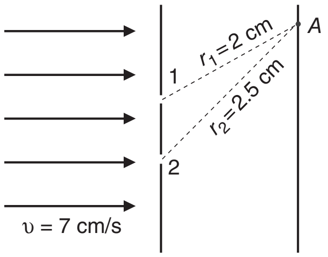
$$\text{(a) } \hbar \to 0, \qquad \text{(b) } m \to \infty\,?$$

First, state your prediction—taking care to explain your rationale—and then check whether it is correct.

1.8 Apply Bohr's quantization condition to obtain the allowed energies of a particle in a central force field, like the one in hydrogen, but with a force law $F = -kr$, known as a three-dimensional harmonic oscillator. Do the same for $F = -gr^3$.

1.9 In a Davisson–Germer experiment—where $d = 2.15\,\text{Å}$ (nickel crystal)—the electrons of the incident beam have energy 64 eV. What is the angle of oblique reflection in this case?

1.10 In the (hypothetical) double-slit experiment of the figure, you are asked to decide whether point A on the screen will be a local maximum or a local minimum of the expected interference pattern. Use the rounded values $h \approx 7 \times 10^{-27}\,\text{erg s}$, $m_e \approx 10^{-27}\,\text{g}$ for simplicity. What happens when the speed of the electrons becomes (a) two times greater and (b) four times greater?



1.4 Dimensional Analysis and Quantum Physics

1.4.1 The Fundamental Theorem and a Simple Application

In its simplest version, dimensional analysis is merely a test of the dimensional correctness of a physics formula (i.e., whether both sides of the formula have the same physical dimension). Let us call this the *passive use* of the dimensional method: Given a formula, we simply check whether it is dimensionally correct. But here we wish to speak of the *active use* of the method: how to use dimensional analysis to find a formula that describes a physical phenomenon without even knowing its underlying theory! The conditions under which we can do this are expressed in the following *fundamental theorem of dimensional analysis*.

Theorem 1.1 *If a physical quantity—say, D —depends only on three others—let us call them A, B, C —then the dependence of D on A, B, C is determined on purely dimensional grounds up to a dimensionless multiplicative constant. In particular, we will have*

$$D = \sigma A^{\alpha} B^{\beta} C^{\gamma}, \tag{1.50}$$

where σ is an arbitrary dimensionless constant and α, β, and γ are suitable exponents that are determined by equating the physical dimensions of length, mass, and time of both sides of the equation.

Let us see how the method works in a simple problem from elementary physics.

Example 1.6 Use dimensional analysis to find the formula for the angular frequency ω of a harmonic oscillation as a function of the parameters of the problem.

Solution: The crucial step appears at the very end of the statement of the problem. On which parameters of the problem do we expect the quantity ω to depend? Two obvious candidates are k and m: the spring constant that determines the restoring force via the known relation $F = -kx$ and the mass m of the oscillating body. Is there a third parameter? Come to think of it, this ought to be the maximum displacement a of the body from its equilibrium position; that is, its oscillation amplitude. At this point, the reader may object that the amplitude does not appear in the familiar formula $\omega = \sqrt{k/m}$. But recall that we are not supposed to know anything about this formula, so we should consider all reasonable options. Based now on this theorem and the dependence of the frequency ω on the three quantities k, m, and a, we have

$$\omega = \sigma k^{\alpha} m^{\beta} a^{\gamma}. \tag{1}$$

To determine α, β, and γ, we start from the expressions

$$[\omega] = T^{-1}, \quad [m] = M, \quad [a] = L, \tag{2}$$

while for k the definition $F = -kx$ yields

$$[k] = \frac{[F]}{L} = [M \cdot \text{acceleration}]L^{-1} = (M \cdot LT^{-2})L^{-1} = MT^{-2}, \tag{3}$$

where the brackets denote the physical dimension of a physical quantity with respect to length, mass, and time—L, M, and T—which are the basic units in the cgs system. By substituting now (2) and (3) into (1) we obtain

$$T^{-1} = (MT^{-2})^{\alpha} M^{\beta} L^{\gamma} = L^{\gamma} M^{\alpha+\beta} T^{-2\alpha}. \tag{4}$$

Upon equating the dimensions of length, mass, and time of both sides in (4), we get

$$\gamma = 0, \quad \alpha + \beta = 0, \quad -2\alpha = -1 \Rightarrow \alpha = 1/2, \quad \beta = -1/2, \quad \gamma = 0,$$

which means that the desired formula has the form

$$\omega = \sigma \sqrt{\frac{k}{m}},$$

where σ is the anticipated dimensionless constant. The latter cannot be determined on dimensional grounds alone, but we can calculate it with *a single experiment* for two convenient values of k and m. It should also be stressed that the independence of ω from the oscillation amplitude a—a very special feature of the harmonic oscillator—emerged through purely dimensional arguments, so it does not depend on the details of the theoretical description of the problem. It is a purely dimensional result.

An equally useful exercise for the readers is to prove, using dimensional arguments again, that for a nonlinear power law of the form $F = -kx^3$, the formula for ω becomes

$$\omega = \sigma\sqrt{\frac{k}{m}}\,a$$

whose characteristic feature is the linear dependence of ω on the amplitude of oscillation. The greater the amplitude, the greater the frequency of oscillation and thus the smaller its period. Can you explain why?

We hope the given example has also clarified the reason the fundamental theorem of dimensional analysis holds. It does so because, upon equating the dimensions L, M, T of both sides of (1.50), we obtain a system of three equations with three unknowns from which the desired exponents α, β, and γ can be uniquely determined. It can also be easily shown that, among all possible functional forms $D = f(A, B, C)$, only the form (1.50)—that is, a product of powers—can be compatible with the requirement that both sides of the equation have the same dimension.[11] Also, it goes without saying that the three quantities A, B, and C in (1.50) are dimensionally independent. That is, none of them can be expressed in terms of the other two. It is equally clear that when the desired quantity D depends, not on three, but on four or more quantities, then the dimensional method cannot by itself determine the desired formula, not even up to a dimensionless multiplicative constant. And yet, even in those cases, a suitable use of the dimensional method can lead to remarkable results, depending on whether certain dimensionless parameters of the problem can be considered small.

Closing, we cannot fail to note how advantageous the cgs system of units is over SI, from the perspective of dimensional analysis. In cgs units, the basic physical quantities are only three—L, M, T—while in SI units we should also add the electric charge as an independent unit. Since this increase of the basic quantities in SI units is completely artificial (as artificial as measuring temperature in nonmechanical units), it can easily be remedied. All it takes is to realize that the two systems differ mainly with respect to Coulomb's law, which—in

---

[11] For example, if the function $f(A, B, C)$ could be expanded in a *Taylor series*—normally this is not possible—then we would have

$$D = f(A, B, C) = \sum_{\mu,\nu,\lambda} \sigma_{\mu\nu\lambda} A^{\mu} B^{\nu} C^{\lambda} \qquad (\lambda, \mu, \nu \text{ are positive integers or zero}).$$

However, given that of all the products of the series, only one can have the correct dimensions, the form (1.50) emerges as the only possible choice.

the hydrogen atom, for example—has the form e^2/r^2 in cgs as opposed to $(1/4\pi\epsilon_0)e^2/r^2$ in SI units. Clearly, the (more elegant) results of the cgs system are converted to their counterparts in the SI system upon the substitution

$$e^2 \to k_C e^2 \qquad (k_C = 1/4\pi\epsilon_0).$$

A practical question remains. Is there a considerable number of physical phenomena that depend only on three parameters so that one can use the dimensional method to predict their behavior? The answer is yes. The most fundamental physical phenomena have in almost all cases few—and quite often only *three*—parameters, because by their nature they relate to the simplest possible manifestations of fundamental laws. Two important examples we discuss here pertain to this category.

1.4.2 Blackbody Radiation Using Dimensional Analysis

The *thermal radiation* of bodies—also known as *blackbody radiation*—is a fundamental physical phenomenon for a simple reason: It is completely independent of the material the radiating body consists of. After all, it is for this reason that all incandescent bodies "look" exactly the same; they are visually indistinguishable. The physical explanation of this remarkable fact lies in the thermal nature of their light. Namely, before being emitted by any hot body, light interacts repeatedly with its material and eventually reaches thermal equilibrium with the body. By the time light is emitted, it is *thermal light*. Therefore, the spectral distribution of its intensity has a universal character, just like the distribution of molecular speeds for a gas in thermal equilibrium with the walls of its container. The experimental data for the thermal radiation of bodies—and the definitions of the quantities needed for its description—are given in Figure 1.15.

As expected, experimental efforts to investigate thermal radiation focused, right from the start, on its two most prominent features: the total intensity (i.e., the surface area below the corresponding curve) and the location of its

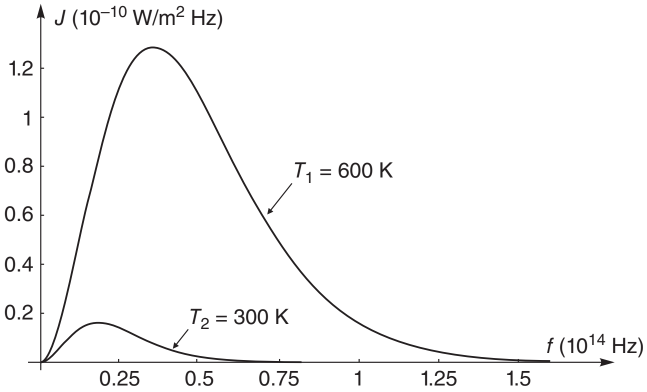


Figure 1.15 The spectral distribution curve of blackbody radiation. J = spectral intensity $\overset{\text{def}}{=}$ radiated energy, per unit time Δt, per unit frequency Δf, and per unit surface ΔS of the radiating body = $\frac{\Delta E}{\Delta t \Delta f \Delta S}$. I = total intensity $\overset{\text{def}}{=} \int_0^\infty J(f,T)df$.

maximum value, as well as the dependence of these features on the temperature of the radiating body. The following empirical laws summarize the major findings on these quantities:

Stefan–Boltzmann law	Wien's law
$I = \sigma T^4$	$\lambda_{\max} = \frac{b}{T}$
The total intensity of the thermal radiation emitted by a blackbody depends on the fourth power of its absolute temperature	The wavelength at which maximum emission occurs is inversely proportional to the absolute temperature of the body
$\sigma = 5.67 \times 10^{-8}\,\mathrm{W/m^2/K^4}$	$b = 0.3\,\mathrm{cm\,K} \Rightarrow \lambda_{\max}(\mathrm{cm}) = \frac{0.3}{T(K)}$

By the end of 1899, Planck had succeeded—utilizing all prior knowledge—to find the full mathematical formula for the spectral distribution $J(f, T)$:

$$J(f, T) = \frac{2\pi h}{c^2} \frac{f^3}{e^{hf/kT} - 1} \qquad \text{(Planck's general empirical formula)}, \quad (1.51)$$

where k is *Boltzmann's constant* and c the speed of light. Immediately thereafter, Planck also concluded that the only theoretical assumption that could explain his formula was the quantization of light. Namely, that the energy of light is quantized in integer multiples of $\epsilon = hf$, where h is our familiar Planck's constant.

In the following examples, we examine what conclusions we can draw on thermal radiation using purely dimensional arguments.

Example 1.7 Use dimensional analysis to make a prediction within the context of classical physics for the spectral distribution $J(f, T)$ of blackbody radiation and comment on the result.

Solution: Our first step is to find the quantities that J depends on. Obviously, two of these are the variables of the problem—the frequency f and temperature T—but also the physical constants of the laws that govern the phenomenon. Such constants are the speed of light c, since we are dealing with an EM phenomenon, and Boltzmann's constant k, since we are also dealing with thermodynamics. On the other hand, J cannot depend on any properties of the light-emitting material because thermal radiation does not depend on these either; it has a universal character. It follows that J cannot depend on any parameters such as the mass and charge of the electron or the nuclear masses, which determine atomic structure and, concomitantly, all properties of macroscopic matter. We thus have

$$J = J(f, T, k, c). \qquad (1)$$

It appears therefore that J depends on four parameters (not three), so the dimensional method is not sufficient to determine the sought dependence. But recall that k is not a truly fundamental constant, but more the result of a historical accident: the fact that we discovered the concept of temperature before figuring out its physical meaning as a measure of the thermal kinetic energy of atoms or molecules in a gas. We thus had to devise, after the fact, the constant k

to ensure that the product kT assigns to T its correct physical meaning and the correct units. It is for this reason that T never appears on its own[12] but always together with k as the product kT.

In this spirit, we can equivalently write (1) as

$$J = J(f, kT, c) \Rightarrow J = \sigma f^{\alpha}(kT)^{\beta}c^{\gamma},$$

with exponents to be calculated in the manner we presented earlier. We thus obtain

$$\alpha = 2, \quad \beta = 1, \quad \gamma = -2 \Rightarrow J \sim f^2 kT/c^2, \tag{2}$$

where we have used the proportionality symbol to avoid a repeated reference to the dimensionless multiplicative constant σ that is always present in formulas derived using dimensional analysis.

Relation (2)—known as the Rayleigh-Jeans law—is a truly profound result. It tells us that applying classical physics to the problem of thermal radiation inescapably leads us to the so-called *ultraviolet catastrophe*: the boundless increase of radiated EM energy at high frequencies. Such an increase would make the total radiated intensity I diverge. So classical physics is not simply unable to explain the phenomenon: It produces an irrational prediction; a *catastrophe*. But there is more. The fact that this prediction resulted solely from dimensional requirements—and not from a detailed calculation that could entail some revisable assumptions—should leave no doubt in our mind that there is really no “cure” for this catastrophe. At least in the problem of thermal radiation, classical physics is fundamentally wrong. And now we know why. In the context of quantum theory, the UV catastrophe is avoided, because at high frequencies the energy hf of the light quanta is so high ($hf \gg kT$) that their thermal excitation is impossible. (Again, quantization saves the day.)

Let us note, finally, that an alternative—and much simpler—way to arrive at relation (2) is to combine the quantities f, kT, and c, to produce the units of the desired quantity J. These units emerge directly from the definition of $J = \Delta E/(\Delta t \cdot \Delta f \cdot \Delta S)$, whence we obtain $[J] = E/L^2$. Of the given quantities, kT has dimensions of energy, while a combination of the other two—c and f—that has dimensions of length is c/f, which, as we know, is the wavelength of the radiation. At this point, we are only interested in the fact that c/f has dimensions of length (i.e., $[c/f] = L$), so the right combination of kT,f and c that yields the correct dimensions E/L^2 for J is $kT/(c/f)^2 = f^2kT/c^2$, as given. For the skeptical readers—to whom this whole process may seem a bit arbitrary—we would stress that, since there is one and only correct combination (according to the theorem), then no matter how we arrived at it, it is bound to be the correct one.

Example 1.8 Use dimensional analysis to see if it can lead us to the abovementioned empirical laws of thermal radiation. That is, the laws of Stefan–Boltzmann and Wien. Assume that we are dealing with a quantum phenomenon, so that Planck’s constant $\hbar$ appears in the formulas, as does the speed of light c.

---

[12] Apart from obvious exceptions (e.g., the specific heat $c_V = (dQ/dT)_V$) where the temperature T, not kT, appears in the definition of the quantity. In this case, all we have to do is a trivial substitution of T with kT and proceed as before.

Solution: The first thing to note is this: Both desired quantities, I and $\lambda_{\max}$, do not depend on f, because I relates to the total radiated power—that is, the integral of J over all f—while $\lambda_{\max}$ relates only to the position of maximum radiation. But I and $\lambda_{\max}$ both depend on the temperature T—always in the form of the product kT—and on c and $\hbar$ as we saw earlier. We thus have

$$I = I(\hbar, c, kT), \quad \lambda_{\max} = \lambda_{\max}(\hbar, c, kT)$$

so the conditions for the fundamental theorem are met, namely, the desired quantity depends only on three physical parameters. Applying the theorem (in the usual systematic manner) yields

$$I \sim \frac{(kT)^4}{c^2\hbar^3}, \quad \lambda_{\max} \sim \frac{\hbar c}{kT}, \tag{1}$$

which are indeed the correct empirical laws—as far as dependence on temperature is concerned—but also with realistic numerical values for the coefficients, as the reader can verify. We can thus confirm the *practical rule* that the dimensionless multiplicative constant in the dimensional method is never a "very large" or a "very small" number. For order-of-magnitude estimates, the undetermined multiplicative constant can safely be regarded as a number of the order of unity.

Note, finally, that formulas (1) can also be derived in the nonsystematic way we sketched earlier. Since $[I] = E/L^2T$, $[\lambda_{\max}] = L$, and kT provides an energy term, we need combinations of kT, c, and $\hbar$ with dimensions of length and time. Such combinations can be obtained easily if we realize that $[\hbar] = ET$ and $[\hbar c] = EL$, so we obtain

$$\left[\frac{\hbar}{kT}\right] = \frac{ET}{E} = T, \quad \left[\frac{\hbar c}{kT}\right] = \frac{EL}{E} = L$$

and then

$$[I] = \frac{E}{L^2T} \Rightarrow I \sim \frac{kT}{\left(\frac{\hbar c}{kT}\right)^2 \cdot \left(\frac{\hbar}{kT}\right)} = \frac{(kT)^4}{c^2\hbar^3}, \quad [\lambda_{\max}] = L \Rightarrow \lambda_{\max} = \frac{\hbar c}{kT}.$$

1.4.3 The Hydrogen Atom Using Dimensional Analysis

It would be interesting to check whether dimensional analysis can tell us something about the two basic empirical quantities regarding the hydrogen atom: the size of the atom—on the order of Å, as we know—and its ionization potential (or energy) whose empirical value is 13.6 eV. We will take it as given that classical physics cannot describe this or any other atom, since the very existence of the atom would not be possible within the framework of classical physics: The electron would have collapsed onto the nucleus. And even if such a collapse were avoided somehow, classical physics cannot provide a mechanism for the atom to have a definite size. Thus, our study will proceed within the quantum context, so that the sought quantities a (atomic radius) and W_{I} (ionization energy) will be functions of the form

$$a = a(\hbar, m, e), \quad W_{\mathrm{I}} = W_{\mathrm{I}}(\hbar, m, e) \tag{1.52}$$

that is, functions of Planck's constant, the electron mass, and the electron charge. Why not the proton mass? Because the proton, being much heavier than the electron, is practically fixed at the center of the atom and its mass can be regarded as infinite to a first approximation. But to the second approximation, the proton mass does play a role, and to account for it we should substitute the electron mass with the *reduced mass* of the electron–proton system. But why would the speed of light not appear in formulas (1.52)? Since the quantum atom—in its ground state—does not radiate, c need not appear and play a role in determining this state. Therefore, the quantities we are interested in depend only on the triplet $\hbar, m$, and e. We thus determine their dependence using dimensional arguments as usual to obtain

$$a \sim \frac{\hbar^2}{me^2}, \quad W_{\mathrm{I}} \sim \frac{me^4}{\hbar^2}. \tag{1.53}$$

Since the combination $e^2/c\hbar$—the fine structure constant—is dimensionless, it is clear that $e^2/\hbar$ has dimensions of speed. So the combination $m(e^2/\hbar)^2 = m \times (\text{speed})^2$ has dimensions of energy and is thus the desired unique expression for W_{I}. Moreover, the ratio e^2/a has energy dimensions (it is the potential energy of two electron charges separated by a distance a); if we equate it with the expression for W_{I}, we obtain the dimensionally correct formula for a. As for the numerical values of the quantities in (1.53), we recall from previous calculations that 0.5 Å is the value for a and 27.2 eV is the value for W_{I}, both in the correct order of magnitude for these quantities. Thus, the mere introduction of Planck's constant, aided by dimensional analysis, can give us plausible results for the hydrogen atom (even without a detailed theory). Surely this is a strong indication that in the correct theory for the atom, Planck's constant will play a key role. Note also that the expression $me^4/\hbar^2$, apart from being correct as an order of magnitude, is also exactly twice the ionization potential. Therefore, we can comfortably assume that the formula

$$W_{\mathrm{I}} = \frac{me^4}{2\hbar^2} \tag{1.54}$$

is *exact* and that the empirical relation for energies, $E_n = -13.6\,\mathrm{eV}/n^2$, can also be written as

$$E_n = -\frac{me^4}{2\hbar^2}\frac{1}{n^2}, \quad (n = 1, 2, \dots, \infty) \tag{1.55}$$

which is indeed the correct mathematical expression for the allowed energies of the atom, as we shall see later.

If you now combine what we said here and our discussion in Section 1.3.9, you will arrive at a plausible explanation of how Bohr was led to his theory. As we shall see again and again in the book, dimensional analysis can be a *powerful tool.*

Problems

1.11 Use dimensional analysis to predict—up to a multiplicative dimensionless constant—the formula for the angular frequency ω of a nonlinear oscillator with a force law of the form $F = -kx^3$. What is the most interesting feature of your result?

1.12 Use dimensional analysis to show that in a world of d dimensions, the formulas for the quantities J_{cl} (classical prediction for the spectral distribution of blackbody radiation), I, and $\lambda_{\max}$ will have the form

(a) $J_{\text{cl}} \sim kT\left(\dfrac{f}{c}\right)^{d-1}$,

(b) $I \sim \dfrac{(kT)^{d+1}}{\hbar^d c^{d-1}}$,

(c) $\lambda_{\max} \sim \dfrac{\hbar c}{kT}$.

1.13 The Stefan–Boltzman law is, of course, expected to be a special consequence of Planck's general formula (1.51) for the spectral distribution of blackbody radiation. Show that this is indeed the case and that the theoretical prediction for σ is

$$\sigma = \frac{2\pi^5}{15}\frac{k^4}{c^2\hbar^3}.$$

Does this prediction agree with the experimental value of σ?

1.14 You suspect that you may have made an error in copying σ (i.e., the Stefan–Boltzmann constant) from some book as $\sigma = 5.67 \times 10^{-4}\,\mathrm{W/m^2/K^4}$. Can you argue, using facts from everyday experience, why this numerical value is completely wrong?

1.15 Consider a hypothetical universe where the value of Planck's constant is 10 times lower compared to ours. Would the radiative intensity of a hot body be different in such a universe compared to ours? If yes, then by how much?

1.16 As the universe expands, the wavelength of cosmic microwave background (CMB) photons—this wonderful thermal afterglow of the Big Bang—gets "stretched" by the same factor. This is because, as space itself expands, the distance between two successive crests (or troughs) of a propagating EM wave increases. Given that the CMB is presently observed to have a temperature of $T \approx 3\,\mathrm{K}$, calculate the following: (a) The present intensity of the CMB and (b) the intensity and peak-emission wavelength of the CMB when the universe was 10 times smaller than its present size (i.e., when the distance between two distant galaxies was 10 times smaller than what is presently observed).

1.17 The energy of the photons corresponding to the peak emission of a hot body is equal to 4 eV. Calculate the total emitted intensity from that body, in units of $\mathrm{W/m^2}$.

Further Problems

1.18 Consider the general case of Compton scattering, whereby the wavelength λ of the impinging photon is $\lambda = k\lambda_{\text{C}}$, where k is an arbitrary dimensionless number, and the scattering angle θ is also arbitrary. Use momentum and energy conservation to calculate the quantities λ', p'_γ, ϵ'_γ (= wavelength,

momentum, and energy of the scattered photon) as well as ϕ, p, and E (= scattering angle, momentum, and energy of the scattered electron). Specifically, show that the following relations hold.
For the photon:

$$\lambda' = \lambda_C(1 + k - \cos\theta), \quad p'_\gamma = \frac{mc}{1 + k - \cos\theta}, \quad \epsilon'_\gamma = \frac{mc^2}{1 + k - \cos\theta}.$$

For the electron:

$$\tan\phi = \frac{k}{1+k}\frac{1}{\tan(\theta/2)}, \quad K \equiv E - mc^2 = mc^2\frac{1 - \cos\theta}{k(1 + k - \cos\theta)},$$

$$p = mc\frac{1 - \cos\theta}{k(1 + k - \cos\theta)}\sqrt{\frac{k^2}{\tan^2(\theta/2)} + (k+1)^2},$$

while for the photon before the "collision" we have

$$\lambda = k\lambda_C, \quad p_\gamma = \frac{mc}{k}, \quad \epsilon_\gamma = \frac{mc^2}{k}.$$

Once you confirm that these general formulas reproduce the results of Example 1.2, apply them to obtain the results in the following two special cases:
(a) $k = 1, \theta = \pi/2$,
(b) $k = 2, \theta = 2\pi/3$.

1.19 Aside from its historical role in the development of quantum theory, Bohr's quantization condition—that only those circular orbits are allowed for which the angular momentum $\ell = mvr$ of the electron is an integer multiple of Planck's constant $\hbar$—is still useful, as it provides a quick way to approximately calculate the allowed energies for various central force fields. In fact, these calculated energies—with a possible exception of the first few of them—reflect the essential features of the energy spectrum. In this spirit, apply Bohr's condition $\ell = n\hbar$ to show that for an attractive central force of the form $F = \pm gr^\nu$ (the sign depends on ν being positive or negative, respectively, while g is assumed positive), the radii of the allowed orbits, the speeds, and the energies of the particle on these orbits are given by the formulas

$$r_n = \frac{(\hbar n)^{2/(\nu+3)}}{(mg)^{(\nu+3)}}, \quad v_n = \frac{1}{m}(mg)^{1/(\nu+3)}(n\hbar)^{(\nu+1)/(\nu+3)}, \tag{1}$$

$$E_n = \pm\frac{g}{2}\frac{\nu+3}{\nu+1}\frac{(\hbar n)^{2(\nu+1)/(\nu+3)}}{(mg)^{(\nu+1)/(\nu+3)}}. \tag{2}$$

Do these formulas reproduce the known results for hydrogen? What about the three-dimensional harmonic oscillator, where $F = -gr$?

1.20 It may sound hard to believe, but the temperature at the surface of the sun—a very distant object—can actually be deduced with reasonable accuracy on the basis of the following observations from everyday experience, in conjunction with the law $I = \sigma T^4$ and the known value for σ ($\sigma = 5.67 \times 10^{-8}\,\mathrm{W/m^2/K^4}$).

Observation #1: A solar panel works! That is, it produces roughly the same work (albeit a bit slower) as an electrical device (e.g., water heater) that consumes power on the order of 2–3 kW, which is typical of all energy-intensive home appliances. Needless to say, the light-collecting surface of a solar panel is on the order of 1 $\mathrm{m^2}$.

Observation #2: The solar disk—whose apparent size is roughly equal to that of the moon's disk—can be blocked out (as you can verify yourselves) by an object (e.g., part of our finger) of size 1 cm at the end of our stretched arm, which is roughly 1 m away from our eyes.

Utilize these two observations to calculate—or, at least, estimate—the temperature at the sun's surface.

1.21 Given that life on earth is the outcome of a primordial evolutionary adaptation to environment, a key element of which is sunlight (a source of both energy and "information" about the world around us), it is not unreasonable to assume that terrestrial living beings gradually "tuned in" to the sun, to utilize the energetically richer region of its spectrum. After all, this is the only way to explain the remarkable fact that the eyes of all living beings "see" roughly in the same spectral range. Actually, the same holds true for all light-collecting molecules (e.g., chlorophyll) of plants.

Use the given reasoning to estimate—in conjunction with the formula $\lambda_{\max}\,(\mathrm{cm}) = 0.3/T(\mathrm{K})$—the temperature at the sun's surface. Does your result roughly agree with what you obtained in the previous problem?

1.22 Use dimensional analysis to predict (without referring to Bohr's theory) the dependence of the radius of the ground state orbit, and the corresponding energy, on the parameters $\hbar$, m, and g for an arbitrary central potential of the form $V(r) = \pm g r^{\nu}$ for a given ν. Do your results agree with those of Problem 1.19?