

3 Scattering: The Life of Photons

Between birth (emission) and death (absorption) photons are scattered. That is, their direction of propagation changes. The zigzag course of a photon as it wends its way through matter is, so to speak, its biography. In this chapter we take a more catholic view of scattering than is the norm. Unification long has been an aim of physics because it enables our limited human minds to comprehend a universe of bewildering complexity. So among our tasks in this chapter is to remove artificial walls that should have been torn down long ago in the cause of intellectual freedom just as the Berlin Wall was torn down in the cause of political freedom.

3.1 Scattering: An Overview

Why is light scattered? No single answer will satisfy everyone, yet because scattering by particles is amenable to treatment mostly by classical electromagnetic theory, our answer lies within this theory.

Although palpable matter may appear continuous and often is electrically neutral, it is composed of discrete electric charges. Light is an oscillating electromagnetic field, which can excite these charges to oscillate. Oscillating charges radiate electromagnetic waves, a fundamental property of such charges with its origins in the finite speed of light. These radiated electromagnetic waves are scattered waves, excited by a source external to the scatterer. Incident waves from the source excite secondary waves from the scatterer, and the superposition of all these waves is what is observed. The secondary waves are said to be *elastically* scattered if their frequency is that of the source (*coherently* scattered also is used).

Scientific knowledge grows like the accumulation of bric-a-brac in a vast and disordered closet in a house kept by a sloven. Few are the attempts at ridding the closet of rusty or broken or obsolete gear, at throwing out redundant equipment, at putting things in order. For example, spurious distinctions still are made between reflection, refraction, scattering, interference, and diffraction despite centuries of accumulated knowledge about the nature of light and matter.

Why do we think of specular (mirror-like) reflection as occurring *at* surfaces rather than *because* of them whereas we usually do not think of scattering by particles in this way? One reason is that we can see and touch the surfaces of mirrors and ponds. Another reason is the dead hand of traditional approaches to the laws of specular reflection and refraction.

The empirical approach arrives at these laws as purely geometrical statements about what is observed, and a discreet silence is maintained about underlying causes (always a safe course). The second approach is by way of continuum electromagnetic theory: reflected and refracted waves satisfy the partial differential equations of the electromagnetic field (the Maxwell equations). Perhaps because this approach, which yields the amplitudes and phases

of waves, entails imposing conditions at boundaries, reflected and refracted waves are mistakenly thought to originate from boundaries rather than from all the matter they enclose. This second approach comes to grips with the nature of light but not of matter, which is treated as continuous. The third approach is to recognize explicitly that reflection and refraction are consequences of scattering of waves by discrete matter. Although this scattering approach was developed by Paul Ewald and Carl Wilhelm Oseen early in the last century, it has diffused with glacial slowness. When the optically smooth interface between optically homogeneous media is illuminated, the reflected and refracted waves are superpositions of vast numbers of secondary waves excited by the incident wave. Moreover, every molecule, not just those at or near the interface, contributes to the total. Thus reflected and refracted light is, at heart, an interference pattern of light scattered by discrete molecules (see Sec. 7.2.1 for more about this). The fourth approach is to recognize the discreteness of both matter and radiation fields. This is the method of quantum electrodynamics, presumably the most rigorous but, alas, nearly impossible to apply except for very simple systems (e.g., the hydrogen atom).

No optics textbook would be complete without sections on interference and diffraction, a distinction without a difference: there is no diffraction without interference. Moreover, diffraction is encumbered with many meanings: a synonym for scattering; small deviations from rectilinear propagation; wave motion in the presence of an obstacle; scattering by a flat obstacle; any departure from geometrical (ray) optics; scattering near the forward direction; and scattering by a periodic array. A term with so many meanings has no meaning. Even the etymology of diffraction is of little help, coming from a Latin root meaning to break, the origin of fraction, fracture, fractal, and fracas.

There is no fundamental difference between diffraction and scattering. Scattering by a sphere (see Sec. 3.5.1) is sometimes called diffraction by a sphere. For many years we have offered a million-dollar prize to anyone who can devise a detector that distinguishes between scattered and diffracted waves, accepting the one but rejecting the other. So far no one has collected, and the money continues to draw interest in a numbered Swiss bank account.

The only meaningful distinction is that between approximate theories. What are called diffraction theories often obtain answers at the expense of obscuring if not completely distorting the physics of the interaction of light with matter. For example, an illuminated slit in an opaque screen may be the mathematical source but it is not the physical source of a so-called diffraction pattern. Only matter in the screen can give rise to secondary waves that superpose to yield the observed pattern. Yet generations of students have been taught that empty space is the source of the radiation from a slit. To befuddle them even more, they also have been taught that two slits give an interference pattern whereas one slit gives a diffraction pattern. But every pattern of scattered light called a diffraction pattern is a consequence of interference. And every theory bearing the label diffraction is a wave theory because only such theories can account for interference.

A variation on the bogus notion that empty space is the source of electromagnetic waves is the oft-repeated mantra that a changing electric field “produces” a magnetic field, and a changing magnetic field “produces” an electric field. Not true. Electric fields are produced by charges, magnetic fields by charges in motion (currents). There are always material *sources* of electromagnetic fields; they do not arise out of empty space.

If we can construct a mathematical theory (called for no compelling reason diffraction theory) that enables us to avoid having to consider the nature of matter, all to the good. But

this theory and its quantitative successes should not blind us to the fact that we are pretending. Sometimes the pretense cannot be maintained, and when this happens a finger is mistakenly pointed at “anomalies”, whereas what is truly anomalous is that a theory so devoid of physical content could ever give adequate results.

A distinction must be made between a physical process and the superficially different theories used to describe it. There is no fundamental difference between specular reflection and refraction by films, diffraction by edges or slits, and scattering by particles. All are consequences of light exciting matter to radiate. The only difference is in how this matter is arranged in space and the approximate theories sufficient for a quantitative description of the scattered light. Different terms for the same physical process are incrustations deposited during the evolution of our understanding of light and its interaction with matter.

3.2 Scattering by a Dipole

The electric charges in matter are acted on by electric fields. Indeed, they are defined as forces per unit charge. When exposed to the oscillatory electric field of an electromagnetic wave, charges in matter are set into oscillation. The fields of these charges therefore change but not instantaneously. This change is propagated outward from the charges in the form of waves traveling at the speed of light. The simplest example of an electrically neutral system is a dipole.

In Section 2.6 we considered the forces acting on a damped oscillator (dipole). We discovered that a dipole, when acted on by a time-harmonic electric field, oscillates at the same frequency as this field. But if the dipole oscillates, it must radiate, and radiation carries energy. So our analysis was incomplete. We didn't account for a dipole losing energy by radiating. This can be accounted for by saying that it is as if the dipole were acted upon by another dissipative force called the *radiative reaction*. Work per unit time done by this force is the rate of energy radiated by the dipole. The radiative reaction is proportional to the second time-derivative of velocity, and hence the equation of motion of a dipole for which the radiative reaction is included is

$$m \frac{d^2 \mathbf{x}}{dt^2} = -K \mathbf{x} - b \frac{d\mathbf{x}}{dt} + w \frac{d^3 \mathbf{x}}{dt^3} + e \mathbf{E}_o \exp(-i\omega t), \quad (3.1)$$

where w is a constant.

As in Section 2.6 we are interested in the steady-state solution to this equation,

$$\mathbf{x} = \mathbf{x}_o \exp(-i\omega t). \quad (3.2)$$

Substituting Eq. (3.2) into Eq. (3.1) yields

$$\mathbf{x}_o = \frac{e}{m_o} \frac{\mathbf{E}_o}{\omega_o^2 - \omega^2 - i\gamma\omega}, \quad (3.3)$$

where

$$\gamma = \gamma_a + \gamma_s \omega^2 / \omega_o^2 \quad (3.4)$$

and

$$\gamma_a = \frac{b}{m}, \quad \gamma_s = \omega_o \frac{w}{m}. \quad (3.5)$$

Equation (3.3) has the same form as Eq. (2.61), so we can immediately write the expression for the time-averaged rate at which work is done (power) on the dipole by the incident electric field, which falls naturally into two parts:

$$\langle P \rangle = \langle P_a \rangle + \langle P_s \rangle, \quad (3.6)$$

where

$$\langle P_a \rangle = \frac{e^2}{2m} E_o^2 \frac{\omega^2 \gamma_a}{(\omega_o^2 - \omega^2)^2 + \gamma^2 \omega^2}, \quad (3.7)$$

$$\langle P_s \rangle = \frac{e^2}{2m} \frac{E_o^2}{\omega_o^2} \frac{\omega^4 \gamma_s}{(\omega_o^2 - \omega^2)^2 + \gamma^2 \omega^2}. \quad (3.8)$$

We interpret $\langle P_a \rangle$ as the power absorbed and $\langle P_s \rangle$ as the power scattered by the dipole. Absorbed power is a consequence of transforming radiant energy into other forms. Scattered power is a consequence of incident radiation in one direction exciting radiation in all directions. $\langle P_s \rangle$ is the time-average of the power radiated (scattered) in all directions by the dipole. As required, $\langle P_a \rangle$ vanishes as γ_a (dissipation) vanishes, as does $\langle P_s \rangle$ when γ_s (radiative reaction) vanishes. Scattering is not a transformation of radiant energy into other forms but rather into different directions; radiant energy is conserved.

In the scattering considered here the frequency of oscillation of the dipole, and hence the frequency of the radiation it scatters, is equal to the frequency of the incident radiation (*elastic scattering*). But all scattering is not elastic. Indeed, there is no such thing as truly elastic scattering if for no other reason than that matter is in motion, so the frequency of the radiation scattered by moving matter is Doppler-shifted (see Sec. 3.4.6) from that of the incident radiation. The term *quasi-elastic scattering*, which sometimes is used, signals that strict elastic scattering is a mathematical fiction.

If the frequency ω of the incident electric field is much less than the resonant frequency ω_o (frequency at the center of the absorption band) and the width of the band γ is much less than ω_o

$$\langle P_s \rangle \approx \frac{e^2}{2m} \frac{E_o^2}{\omega_o^6} \gamma_s^2 \omega^4. \quad (3.9)$$

This is the much-celebrated scattering law first derived by Lord Rayleigh in 1871. Frequency to the fourth power is proportional to wavelength to the inverse fourth power. We have more to say about the observable consequences of this law in Section 8.1.

We obtained Rayleigh's scattering law in the following form: the radiation scattered in all directions by a dipole, for a given irradiance (proportional to E_o^2), is proportional to the fourth power of the frequency of the incident radiation that excites this scattered radiation. But we also recognize that Rayleigh's law is an approximation, not valid at frequencies near an

absorption band. That is, strict ω^4 scattering by a dipole is yet another mathematical fiction, a useful approximation provided we recognize its limitations.

Equations (3.7) and (3.8) also show that scattering and absorption are not completely independent of each other. Absorbed power depends on γ , which has a contribution from scattering (γ_s) and absorption (γ_a), and similarly for scattered power.

Just as we defined the absorption cross section in Section 2.4 we can define the scattering cross section of a molecule (dipole) σ_s by

$$\langle P_s \rangle = W_s = \sigma_s F, \quad (3.10)$$

where W_s is the radiant power scattered in all directions by the molecule and F is the irradiance of the beam that excites this scattered radiation. Unlike the absorption cross section, the scattering cross section can be apportioned into contributions in different directions. No scatterer of electromagnetic radiation scatters exactly the same in all directions (Sec. 7.3), and scattering can be highly directionally asymmetric depending on the size of the scatterer relative to the wavelength (Sec. 3.5.3).

We use the term dipole and molecule almost as synonyms because for our purposes every molecule is a dipole, although not conversely. If sufficiently small, a particle can be a dipole oscillator. To explain why requires a firm understanding of interference, which in turn requires a firm understanding of waves, which we turn to next.

3.3 Waves on a String: The One-Dimensional Wave Equation

The easiest path to understanding waves is by way of the simplest example: waves on a string under tension (Fig. 3.1). In its equilibrium position the string lies along the x -axis, but if the string is displaced in the y -direction it will vibrate because of opposition between the inertia of the string and a restoring force provided by the tension. Consider a segment of string lying between x and $x + \Delta x$. If σ is the mass per unit length of the string (assumed constant), the equation of motion of this segment is

$$\sigma \Delta x \frac{\partial^2 y}{\partial t^2} = F_y, \quad (3.11)$$

where F_y is the y -component of the force on Δx , a consequence of the tension, assumed constant. This assumption requires that the lateral displacement of the string be small. We also assume that the string always lies in the xy -plane. The total force acting on Δx by the string on both sides of it is

$$F_y = T \sin\{\vartheta(x + \Delta x)\} - T \sin\{\vartheta(x)\}, \quad (3.12)$$

where T is the tension in the string and ϑ is the angle between the string and the x -axis. Unless the tension is very small, the force of gravity is negligible. Because of the assumption of small displacements

$$\sin \vartheta \approx \tan \vartheta = \frac{\partial y}{\partial x}. \quad (3.13)$$

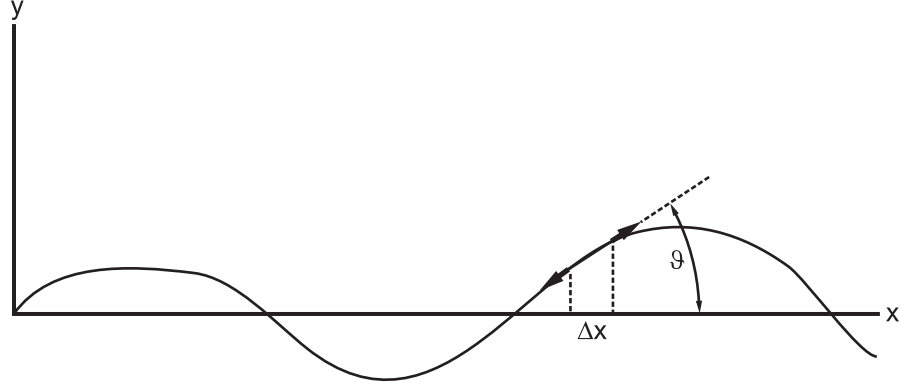


Figure 3.1: The (small) y -displacement of a uniformly tense string stretched along the x -axis varies in space and time. The angle ϑ between the tangent to the string at any point and the x -axis is shown greatly exaggerated.

If Eqs. (3.11)–(3.13) are combined we obtain

$$\sigma \Delta x \frac{\partial^2 y}{\partial t^2} = T \left(\frac{\partial y}{\partial x} \right)_{x+\Delta x} - T \left(\frac{\partial y}{\partial x} \right)_x. \quad (3.14)$$

Divide both sides of this equation by Δx and take the limit as $\Delta x \rightarrow 0$:

$$\sigma \frac{\partial^2 y}{\partial t^2} = T \frac{\partial^2 y}{\partial x^2}. \quad (3.15)$$

We also can write this equation as

$$\frac{1}{v^2} \frac{\partial^2 y}{\partial t^2} = \frac{\partial^2 y}{\partial x^2}, \quad (3.16)$$

where

$$v = \sqrt{T/\sigma}. \quad (3.17)$$

Note that v has the dimensions of speed, and hence is called the *phase speed* (about which we say more in this section and Sec. 3.5.1). Equation (3.16) is a *wave equation*, specifically, a one-dimensional scalar wave equation. A wave is anything that satisfies a wave equation. If you know something about one kind of wave you know something about all kinds of waves. Begin with the simplest kind, understand it, and you are on your way to understanding more complicated waves (e.g., three-dimensional vector waves).

Multiply both sides of Eq. (3.15) by the time derivative of y :

$$\sigma \frac{\partial y}{\partial t} \frac{\partial^2 y}{\partial t^2} = \frac{\partial}{\partial t} \left\{ \frac{1}{2} \sigma \left(\frac{\partial y}{\partial t} \right)^2 \right\} = T \frac{\partial y}{\partial t} \frac{\partial^2 y}{\partial x^2}. \quad (3.18)$$

The left side of Eq. (3.18) is the time rate of change of the kinetic energy (per unit length) of the string. This is a clue that we are headed toward an energetic form of the equation

of motion. Before obtaining this equation we pause to reflect that although multiplying the equation of motion by partial derivatives and rearranging terms cannot change its physical content, it can result in a more felicitous form. To proceed we note that

$$\frac{\partial}{\partial x} \left(\frac{\partial y}{\partial x} \frac{\partial y}{\partial t} \right) = \frac{\partial y}{\partial t} \frac{\partial^2 y}{\partial x^2} + \frac{\partial^2 y}{\partial x \partial t} \frac{\partial y}{\partial x} \quad (3.19)$$

and

$$\frac{\partial}{\partial t} \left(\frac{\partial y}{\partial x} \right)^2 = 2 \frac{\partial y}{\partial x} \frac{\partial^2 y}{\partial t \partial x}. \quad (3.20)$$

With the assumption that the order of differentiation is irrelevant,

$$\frac{\partial^2 y}{\partial x \partial t} = \frac{\partial^2 y}{\partial t \partial x}, \quad (3.21)$$

Equations (3.18)–(3.21) can be combined to yield

$$\frac{\partial}{\partial t} \left\{ \frac{1}{2} \sigma \left(\frac{\partial y}{\partial t} \right)^2 + \frac{1}{2} T \left(\frac{\partial y}{\partial x} \right)^2 \right\} = T \frac{\partial}{\partial x} \left(\frac{\partial y}{\partial t} \frac{\partial y}{\partial x} \right). \quad (3.22)$$

This equation resembles energy conservation equations of the kind encountered in the dynamics of systems of point particles. The only difference is that the system of interest here is continuous rather than discrete, as evidenced by the partial derivatives. By now we should suspect that the second term in brackets on the left side of Eq. (3.22) is some kind of potential energy per unit length of string.

A short segment of string with x -coordinates x and $x + \Delta x$ has length Δx when the string lies along the x -axis. In general, the length of this segment is Δs , where

$$\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2} = \Delta x \sqrt{1 + \left(\frac{\Delta y}{\Delta x} \right)^2}. \quad (3.23)$$

If the slope of the string is everywhere much less than unity [an assumption made previously in Eq. (3.13)], the change in length of the segment is

$$\Delta s - \Delta x \approx \Delta x \left\{ 1 + \frac{1}{2} \left(\frac{\Delta y}{\Delta x} \right)^2 \right\} - \Delta x. \quad (3.24)$$

Divide both sides of this equation by Δx and take the limit as $\Delta x \rightarrow 0$ to obtain the change in length per unit length

$$\frac{1}{2} \left(\frac{\partial y}{\partial x} \right)^2. \quad (3.25)$$

This quantity multiplied by the tension T is the work done (per unit length) in stretching the string, and is the second term in brackets on the left side of Eq. (3.22). Hence we are justified in calling this term a potential energy (per unit length).

If we integrate Eq. (3.22) from x_1 to x_2 we obtain

$$\frac{d}{dt} \int_{x_1}^{x_2} \left\{ \frac{1}{2} \sigma \left(\frac{\partial y}{\partial t} \right)^2 + \frac{1}{2} T \left(\frac{\partial y}{\partial x} \right)^2 \right\} dx = T \left(\frac{\partial y}{\partial x} \frac{\partial y}{\partial t} \right)_2 - T \left(\frac{\partial y}{\partial x} \frac{\partial y}{\partial t} \right)_1. \quad (3.26)$$

The integral is the total energy (kinetic plus potential) of the string between x_1 and x_2 . Thus the time rate of change of the energy of this system is equal to a quantity evaluated at one boundary of the string less that same quantity evaluated at the other boundary. It is therefore natural, if not inescapable, to call this quantity an *energy flux*, the rate at which energy is transported across the boundary of the system from its surroundings (the rest of the string). If we define the energy flux vector by

$$T \frac{\partial y}{\partial x} \frac{\partial y}{\partial t} \mathbf{e}_x, \quad (3.27)$$

where \mathbf{e}_x is a unit vector in the positive x -direction, the energy flux into the system is the scalar product of this vector with the outward normals to the system on its boundary, which is just the two points x_1 and x_2 .

3.3.1 Solutions to the Wave Equation

Any twice-differentiable function $f(u)$ of the single variable u can be transformed into a solution to Eq. (3.16) by setting

$$u = x \pm vt. \quad (3.28)$$

That the function of two variables so obtained is a solution is readily verified by repeated use of the chain rule for differentiation

$$\frac{\partial}{\partial x} f\{u(x, t)\} = \frac{df}{du} \frac{\partial u}{\partial x}, \quad (3.29)$$

and so on. Note that f is arbitrary subject only to the requirement that it be twice differentiable, and hence there are infinitely many possible solutions to the one-dimensional, scalar wave equation.

The function $f(x - vt)$ has a simple geometrical interpretation, which helps us to visualize, and hence understand waves. At $t = 0$, $y = f(x)$ describes the shape (displacement) of the string. At some later time Δt , the shape of the string will be different but yet the same because the argument of f , and hence its value, will be the same at $x + \Delta x$ if

$$x = x + \Delta x - v\Delta t, \quad (3.30)$$

from which

$$\frac{\Delta x}{\Delta t} = v. \quad (3.31)$$

Thus the waveform $f(x)$ can be looked upon as propagating (without change of shape) in the positive x -direction with constant speed v .

The energy flux vector corresponding to this wave is

$$-Tv \left(\frac{df}{du} \right)^2 \mathbf{e}_x. \quad (3.32)$$

At the left boundary of the system, the outward normal is $-\mathbf{e}_x$, and hence the energy flux into the system from left to right is positive, as required of a wave propagating in the positive x -direction. At the right boundary, the wave transports energy out of the system.

The function $f(x + vt)$ corresponds to a wave propagating in the negative x -direction. Because the wave equation is linear, a sum of solutions is also a solution. That is, if y_1 and y_2 are solutions, so is $a_1y_1 + a_2y_2$, where a_1 and a_2 are arbitrary constants. Thus it is possible for waves to propagate on a string in both directions simultaneously. Indeed, two equal but opposite (in sign) pulses propagating in opposite directions can meet, annihilate each other briefly when they exactly overlap, then proceed unscathed carrying no memories of or scars from their meeting.

3.3.2 Sinusoidal Wave Functions

One of the simplest possible solutions to the one-dimensional wave equation is a sinusoid:

$$y = a \cos(kx - \omega t), \quad (3.33)$$

where the wavenumber $k = 2\pi/\lambda$ and frequency ω are arbitrary subject to the requirement that $v = \omega/k$. We noted previously (Sec. 2.7.1) that wavenumber could mean $1/\lambda$ or $2\pi/\lambda$; and frequency, unqualified, could be ν or $\omega = 2\pi\nu$. The argument of the cosine, $kx - \omega t$, is called the *phase* of the wave, and Eq. (3.33) describes what is often called a *plane harmonic* wave. Plane because the *surfaces of constant phase*

$$kx - \omega t = \text{const.} \quad (3.34)$$

are planes, harmonic because the time dependence is sinusoidal (harmonious). A harmonic wave is sometimes referred to as monochromatic even when it is not an electromagnetic wave at visible frequencies.

According to Eq. (3.34) the phase at (x, t) will be the same at $(x + \Delta x, t + \Delta t)$ if $\Delta x/\Delta t = \omega/k = v$, hence the name phase speed for v , the speed of surfaces of constant phase (planes perpendicular to the x -axis). This is not, however, the speed of any physical object but rather that of mathematical surfaces.

From Eqs. (3.27) and (3.33) it follows that the (instantaneous) power transmitted by the plane harmonic wave at a left boundary of the string is

$$T\omega ka^2 \sin^2(kx - \omega t). \quad (3.35)$$

Suppose that ω is so high that an energy flux detector responds not to instantaneous but to time averages of Eq. (3.35):

$$T\omega ka^2 \langle \sin^2(kx - \omega t) \rangle, \quad (3.36)$$

where

$$\langle \sin^2(kx - \omega t) \rangle = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \sin^2(kx - \omega t) dt. \quad (3.37)$$

Take the time interval $t_2 - t_1$ to be large compared with $1/\omega$. Transforming the variable of integration to $t' = t + kx/\omega$ is equivalent to setting $x = 0$ in Eq. (3.37). The integrals of $\sin^2 \omega t$ and $\cos^2 \omega t$ over many periods are approximately equal. Because $\sin^2 \omega t + \cos^2 \omega t = 1$ the sum of these two integrals is 1. Thus the integral Eq. (3.37) is approximately $1/2$ and the time-averaged energy flux is

$$\frac{1}{2} T \omega k a^2. \quad (3.38)$$

Because the wave equation is linear, we can deal with complex representations of real wave functions (see Sec. 2.5). That is, the wave function in Eq. (3.33) is the real part of the complex wave function

$$y_c = a \exp(ikx - i\omega t). \quad (3.39)$$

The time-averaged energy flux associated with this wave is therefore

$$\frac{1}{2} T \omega k |y_c|^2. \quad (3.40)$$

3.4 Superposition and Interference

As we noted previously, because of the linearity of the wave equation, the sum (superposition) of solutions is also a solution. Superposition of waves leads to the possibility of *interference*, by which we mean the following. Suppose that plane waves with the same frequency but differing in phase are superposed somehow. The complex representations of these waves are

$$\psi_1 = a_1 \exp(ikx - i\omega t + i\varphi_1), \quad \psi_2 = a_2 \exp(ikx - i\omega t + i\varphi_2), \quad (3.41)$$

where a_1 and a_2 are real constants. The time-averaged power (energy flux) transmitted by the sum of these two waves is (omitting a constant factor), from Eq. (3.40),

$$a_1^2 + a_2^2 + 2a_1 a_2 \cos \Delta\varphi, \quad (3.42)$$

where $\Delta\varphi = \varphi_2 - \varphi_1$ is the phase difference between the two waves, a_1^2 is the power that would be transmitted by ψ_1 alone, and a_2^2 the power that would be transmitted by ψ_2 alone. The additional term is the *interference* term. The cosine of the phase difference (or phase shift) lies between -1 and 1 , and hence the transmitted power lies between $(a_1 - a_2)^2$ and $(a_1 + a_2)^2$. For simplicity assume that the two waves have equal amplitude a ; with this assumption the transmitted power lies between 0 and $4a^2$. The first possibility is often called *destructive interference* and the second *constructive interference*. Destructive and constructive interference do not exhaust the gamut of possibilities: these two terms apply to extremes (and for waves of equal amplitude at that). Depending on the phase difference, the power transmitted by two interfering waves of equal amplitude may be zero or four times the power transmitted by each separately or anything in between. This peculiar property of wave interference is not shared by particles. For example, the kinetic energy transmitted by two billiard balls is the sum of the kinetic energies transmitted by each acting separately.

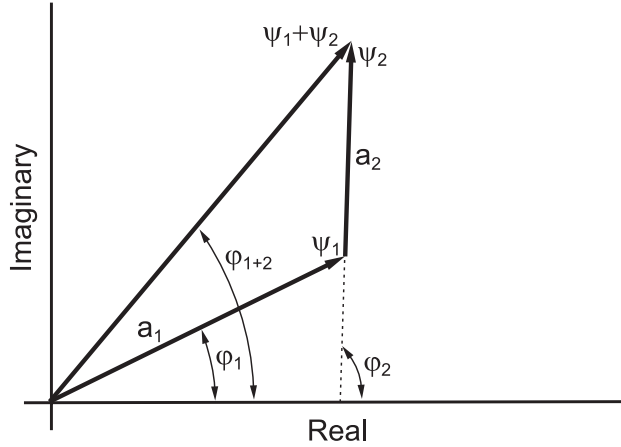


Figure 3.2: Addition of two complex numbers depicted graphically as the addition of two vectors in a plane (the complex plane).

Destructive interference should not be confused with *annihilation* of radiation, nor constructive interference with *creation* of radiation. Equation (3.42) could be interpreted (incorrectly) as implying that two beams can be superposed to create or annihilate radiant energy. But this equation applies strictly to the superposition of two plane waves, and such waves do not exist. If interference results in an increase in radiant power in some directions, this must be compensated for by a decrease in other directions. Radiation can be annihilated (absorbed) or created (emitted) only by matter.

The superposition and interference of waves can be depicted graphically as addition of vectors in the complex plane (see Sec. 2.5). For example, the two waves in Eq. (3.41), omitting the common factor $\exp(ikx - i\omega t)$, are so depicted in Fig. 3.2. A glance at this figure conveys that the resultant of two waves depends on their phase difference. Waves and other time-harmonic quantities represented by vectors in the complex plane are the *phasors* much beloved of electrical engineers (if you have to talk to them – and sometimes this is unavoidable – it helps to know their lingo). The geometrical depiction of waves as phasors helps us understand interference of waves. If the time-harmonic factor $\exp(-i\omega t)$ is included, we have a dynamical picture of a wave as a phasor rotating in the complex plane with constant angular speed ω . Having exposed some of the jargon of electrical engineers we further note that they often use the positive time-harmonic convention and with j in place of i . This is because they fear that their bread-and-butter, electric current, usually represented as i , might be confused with the square root of -1 . Anyone who writes time-harmonic quantities as $\exp(j\omega t)$ is almost certain to be an electrical engineer. By their symbols and conventions ye shall know them.

An essential requirement for interference is *coherence*, a concept so fundamental that we devote a section to it. But first we consider superposition of waves with different frequencies.

3.4.1 Superposition of Waves with Different Frequencies

Two time-harmonic waves with different frequencies are separately time-harmonic but their sum is not (yet another variation on the theme that the sum of exponentials is not an exponential; see Secs. 2.1.3 and 2.3.1). Consider two plane harmonic waves on a string propagating in the positive x -direction with the same phase velocity but different frequencies and wavenumbers:

$$A_1 \cos(k_1 x - \omega_1 t), \quad A_2 \cos(k_2 x - \omega_2 t). \quad (3.43)$$

For simplicity we determine the power transmitted by the sum of these two waves at $x = 0$. The result, after some manipulation, is

$$C_1 \sin^2 \omega_1 t + C_2 \sin^2 \omega_2 t + C_3 \cos(\omega_1 - \omega_2)t + C_4 \cos(\omega_1 + \omega_2)t, \quad (3.44)$$

where C_j is a constant. To derive this equation we used the identity

$$\sin x \sin y = \frac{1}{2} \cos(x - y) - \frac{1}{2} \cos(x + y). \quad (3.45)$$

Equation (3.44) is the instantaneous (not the time-averaged) power, the sum of four terms: two with the frequencies of the separate waves, one with a frequency equal to the difference of frequencies and one with a frequency equal to their sum. When we take the time average, the terms oscillating with the sum and difference of frequencies vanish (provided that the period of the lower frequency term is small compared with the averaging time). As far as time-averaged power transmission is concerned, therefore, two waves of different frequencies do not interfere: the total transmitted power is the sum of the powers transmitted by each wave as if it alone were propagating.

This result underlies the description of polychromatic light as a mixture of waves with different frequencies: the total transmitted power is the sum of transmitted powers for each component of the mixture. Denote by $\Delta\omega$ the difference in the frequencies of two superposed waves. For the time average of the third term in Eq. (3.44) to be negligible requires the averaging time to be large compared with $1/\Delta\omega$. At the frequencies of visible and near-visible light, $1/\omega$ is of order 10^{-14} s. Thus even if $\Delta\omega$ is as small as 10^{-9} , the averaging time need be greater than only about 10 μ s.

3.4.2 Coherence

Interference is possible only because of *coherence*. To cohere means to stick together, usually two or more similar objects. Two waves are said to be coherent, for example, if there is a definite and fixed phase relation between them. In a sense, the waves stick together. Equation (3.42) is a coherent superposition of two waves if the phase difference $\Delta\varphi$ is fixed. Now suppose that this phase difference varies in time much more slowly than the (instantaneous) amplitude of either wave but much more rapidly than the time response of a detector. If $\Delta\varphi$ ranges at least over all values from $-\pi$ to π , the time average of $\cos(\Delta\varphi)$ is zero. The two waves are said to be *incoherent*, and the total power transmitted by them is the sum of the individual transmitted powers.

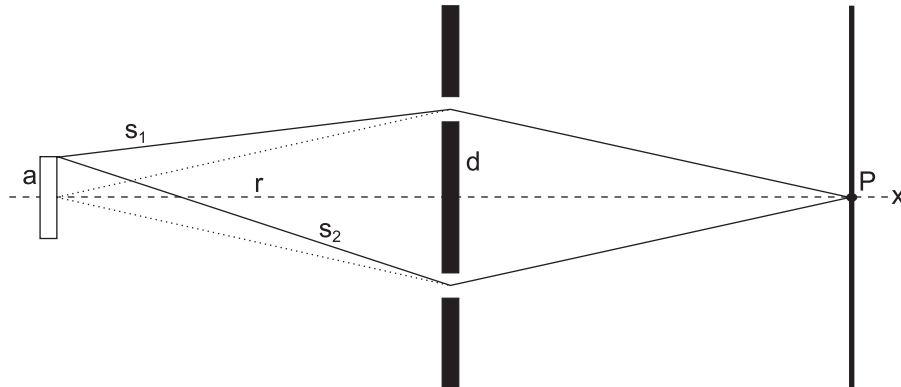


Figure 3.3: An opaque screen with two holes in it separated by a distance $2d$ illuminated by an extended source (disc) of radius a . Each point on the disc, which radiates independently of every other point, is imagined to excite two waves, one from each hole (in reality, it is the screen that radiates). These two waves combine at P , the phase difference between them determined by the path difference $s_2 - s_1$.

Coherent waves must arise from coherent sources. An array of sources is coherent if all its elements bear a fixed spatial relationship to each other. The elements may move and still be coherent, as long as they move in unison. Only a coherent array of sources can give rise to an interference pattern, by which is meant a distribution of radiant power that is not the sum of the radiant powers from each source. Two identical flashlights, for example, do not yield an interference pattern. Shine them onto the same spot on a wall and this illuminated spot is twice as bright as it would be with only one flashlight; the flashlights are mutually incoherent.

Suppose that we illuminate a coherent array of scatterers. Is this sufficient to give an interference pattern? Not unless the array is illuminated by a source of light that itself is coherent to some degree.

Sunlight is sometimes said to be incoherent. Laser light is often said to be coherent. But complete incoherence and coherence are extremes never realized in nature. Every light source is coherent to a degree, and even lasers are incoherent to a degree. One way of specifying the degree of coherence of a light source is by its *lateral coherence length*. To determine this length as simply as possible, consider the two-slit interference problem found in elementary physics textbooks (Fig. 3.3). An opaque screen with two slits (holes) in it is illuminated on one side of the screen, and an observing screen is placed on the other side. Some textbooks show another screen with a single slit or hole between the source (e.g., a light bulb) and the opaque screen with the slits, although the purpose of this mysterious single hole may not be explained. The screen with the two holes in it is a coherent array because the holes are fixed and do not move.

Consider an extended source of light, an incandescent disc, for example, of radius a centered on the axis of the system. The distance between the screen and this source is r . Every point on the source radiates independently of every other point. Waves originating from different points bear no fixed and definite phase relationship. And yet an interference pattern still

is possible if the source is sufficiently small. The holes are separated by a distance $2d$ and the center of the coordinate system lies equidistant between them; the coordinates of the two holes are $(0, d, 0)$ and $(0, -d, 0)$. The source (disc) is centered on the x -axis.

Consider first the source point $(-r, 0, 0)$. In the usual textbook treatment of two-slit interference (criticized in Sec. 3.1), a wave emitted by a source point illuminates the two holes, which magically become sources of outgoing waves (despite the embarrassing fact that this is physically impossible). These two waves combine (interfere) at observation points on the observing screen. The radiant power at these points depends on the phase shift between the two waves, which in turn depends on the different paths they traversed. For simplicity take the observation point P to lie on the axis. With this assumption the difference in paths from the holes to P is zero. The distance from the source point $(-r, 0, 0)$ at the center of the disc to either hole in the screen is the same. Thus the two waves that interfere at P are exactly in phase (no phase shift).

But now consider a source point at the edge of the disc $(-r, a, 0)$. The distance from this point to the nearest hole is

$$s_1 = \sqrt{r^2 + (d - a)^2} \quad (3.46)$$

and the distance to the farthest hole is

$$s_2 = \sqrt{r^2 + (d + a)^2}. \quad (3.47)$$

If d and a are much less than r , the path difference is approximately

$$\Delta s = s_2 - s_1 \approx \frac{2ad}{r}, \quad (3.48)$$

which corresponds to a phase difference [see Eq. (3.33)]

$$\Delta\varphi = 2\pi \left(\frac{\Delta s}{\lambda} \right) = \frac{2\pi}{\lambda} \frac{2ad}{r}. \quad (3.49)$$

This phase difference, the greatest possible, is for points lying on the rim of the incandescent disc. The phase difference is zero for the source point at the center of the disc. Hence all phase differences lie between 0 and $\Delta\varphi$.

If the maximum phase difference is much less than 2π , all points on the source will give essentially the same interference pattern at the observation point. This condition will certainly be satisfied if

$$d \ll \frac{\lambda r}{2a}. \quad (3.50)$$

The solid angle Ω (see Sec. 4.1.1) subtended by the source at the holes is $\pi a^2/r^2$, and hence Eq. (3.50) can be written

$$d \ll \frac{\lambda}{\sqrt{\Omega}}. \quad (3.51)$$

The quantity $\lambda/\sqrt{\Omega}$, called the *lateral coherence length*, is a property of the source. To determine if illumination of a coherent array can yield an interference pattern we need to know the linear dimensions of the array relative to the coherence length. The condition in Eq. (3.51) is a bit strong. We still can get an interference pattern even if d is only less than the coherence length. The mystery of the screen with a single hole placed between an extended source and a screen with two slits now disappears. The function of this first hole is to reduce the angular spread of the light illuminating the slits. Nowadays the two slits are likely to be illuminated by lasers, but before they existed interference patterns were possible using ordinary incandescent light bulbs illuminating a small hole to provide a source with a lateral coherence length larger than the slit spacing.

Note that all points of the source are mutually incoherent: the wave from one point is completely uncorrelated with waves from all other points. Yet this source can still yield interference patterns. Waves from different points do not interfere with each other but rather with themselves (by way of the secondary waves they excite in the slit screen), and if the coherence length is large compared with the slit spacing, each of these separate interference patterns is approximately the same.

At visible wavelengths the coherence length of sunlight is about $50\text{ }\mu\text{m}$, and hence sunlight can give rise to interference patterns. One example is the *corona* (Sec. 8.4.1). The coherent array that produces this pattern is a single cloud droplet, with dimensions of order $10\text{ }\mu\text{m}$, less than the coherence length of sunlight.

Although a single cloud droplet can produce an interference pattern, an array of droplets cannot. That is, waves from different droplets do not interfere. A typical number density of droplets in a cloud is 200 cm^{-3} , which corresponds to an average distance between droplets of about $2000\text{ }\mu\text{m}$, much greater than the coherence length of sunlight. Even if cloud droplets were fixed in space (or at least fixed relative to each other), the waves from different droplets would not interfere, and the power scattered by N droplets still would be N times the power scattered by one.

A striking demonstration of how the coherence properties of an array of scatterers and that of the light illuminating them combine to yield different patterns can be done using a laser with a beam spreader.

An ordinary sheet of white paper in sunlight appears uniformly bright, but that same paper illuminated by a laser exhibits a speckle pattern (*laser speckle*), a mottled pattern of dark and bright regions. The paper is a coherent array: all the fibers in it are fixed in place. And the coherence length of the laser is huge compared with the separation between fibers. Now illuminate a glass of milk with this same laser. The speckle pattern disappears: the milk is uniformly bright. Milk is a suspension of fat globules, the average separation between which is much smaller than the lateral coherence length of the laser. But the globules are in constant motion, and hence do not form a coherent array. Trying to obtain an interference pattern with a suspension of jiggling fat globules would be like trying to obtain a two-slit interference pattern with slits moving rapidly and randomly toward and away from each other.

3.4.3 Distinction between a Theory and an Equation

We criticized the simple theory of the two-slit interference pattern which is based on the physically incorrect assumption that empty space (slits) is a source of waves. Why, then, does

this theory give correct results (sometimes)? Consider first an opaque screen with no slits and illuminated on one side. No light is transmitted. Why? If you accept the superposition principle for electromagnetic waves, you cannot believe that the incident wave is destroyed. It exists everywhere just as it did without the screen in place. But the screen gives rise to secondary waves excited by the primary (incident) wave, and the superposition (interference) of all these waves is what is observed. With the screen in place, interference is destructive everywhere behind it: no *net* wave is transmitted. If it bothers you that the incident wave still exists in the space on the dark side of the screen, consider a standard problem in electrostatics. A conducting shell is placed in an external electric field. Inside the shell the total field is zero (electric shielding). Why? The external field induces a charge distribution on the outer surface of the shell (with zero total charge), within which the total field is the sum of the field of this charge distribution *and* the external field. These two fields are equal and opposite inside the shell so their sum is zero. But no one, to our knowledge, claims that the external field ceases to exist within the shell. To do so would contradict the superposition principle. What is true for electrostatics must also be true for electrodynamics.

An important distinction here (and throughout science), but not made often enough, is that between a theory and an equation. They are not the same. A theory of the two-slit interference pattern based on the literal assumption that empty space is the source of waves is incorrect, but the resulting equation (sometimes) is correct. Who cares as long as “you get the right answer”? Sometimes you *don’t* get the right answer, and if your theory is wrong and you don’t know why you won’t be able to fix it.

Suppose that we remove material from the screen by punching two holes in it. If the interference pattern without the holes results in destructive interference everywhere on the dark side of the screen because of secondary waves excited in the screen material, removing a bit of this material ought to change the pattern. And this is what happens: light is observed at points on the dark side of the screen where previously there was none. To good approximation, it is *as if* the holes themselves were the mathematical sources of waves, whereas their physical source is the entire screen with two holes in it. This approximation cannot be universally valid because it is insensitive to the screen thickness and material and polarization state (Ch. 7) of the incident light. A diffraction grating is, in essence, a periodic array of many slits, and it has been long known that the elementary theory of the grating, which is fundamentally no different from two-slit theory, fails to account for what are often called Rayleigh–Wood anomalies.

In the photon language our description of illumination of an opaque screen is quite different. We would say that incident *photons* do not exist on the dark side of the screen, which does not contradict the assertion that the incident *wave* does. Where confusion may arise is when we ask if there is any incident *light* on the dark side. The answer depends on the language we choose, and hence what we mean by light. And this answer becomes garbled when we speak and think in both languages simultaneously (possibly without realizing it). If it seems enigmatic that an incident wave *can* penetrate solid matter, it is just as enigmatic that incident photons *cannot*. The size of a nucleus, where most of an atom’s mass is concentrated, is smaller than the atom by at least a factor of 10^{-4} , which in turn means that the (volume) fraction of an atom occupied by mass is less than 10^{-12} . Despite its apparent solidity, matter is almost entirely empty space very sparsely populated by charges (electrons and protons). Thus it is not difficult to understand – or at least accept – why waves can penetrate such nearly empty space but not why photons cannot.

3.4.4 Scalar Waves in Three Dimensions

To proceed further we must consider waves in more than one dimension. A generalization of the one-dimensional wave equation [Eq. (3.16)] is

$$\nabla^2 \Psi = \frac{1}{v^2} \frac{\partial^2 \Psi}{\partial t^2}, \quad (3.52)$$

where Ψ is some (unspecified) physical quantity of interest. Equation (3.52) is the governing equation of acoustic waves in a fluid, for example.

The simplest three-dimensional solution to this equation is a plane harmonic wave

$$\Psi = a \exp(i\mathbf{k} \cdot \mathbf{x} - i\omega t), \quad (3.53)$$

where the *wavevector* \mathbf{k} is arbitrary subject to the requirement that its magnitude k satisfy $\omega^2/k^2 = v^2$. That Eq. (3.53) is a solution to Eq. (3.52) can be verified by direct substitution. The surfaces of constant phase for the wave in Eq. (3.53), specified by

$$\mathbf{k} \cdot \mathbf{x} - \omega t = k_x x + k_y y + k_z z - \omega t = \text{const.}, \quad (3.54)$$

are planes to which the wavevector is perpendicular. Thus we can visualize Eq. (3.54) as planes of constant phase propagating in the direction \mathbf{k} with phase speed v .

Any time-harmonic function $\Psi(\mathbf{x}, t) = a(\mathbf{x}) \exp(-i\omega t)$ is a solution to Eq. (3.52) provided that

$$\nabla^2 a + k^2 a = 0. \quad (3.55)$$

If, for example, a varies with the radial coordinate r only, Eq. (3.55) becomes

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{da}{dr} \right) + k^2 a = 0. \quad (3.56)$$

Given Eq. (3.56), the function $u = ar$ satisfies

$$\frac{d^2 u}{dr^2} + k^2 u = 0, \quad (3.57)$$

solutions to which are

$$u = A \exp(\pm ikr), \quad (3.58)$$

and hence

$$\Psi = \frac{A}{r} \exp(\pm ikr - i\omega t) \quad (3.59)$$

is a solution to Eq. (3.52). Equation (3.59) describes *spherical waves*, outgoing (from the origin of coordinates) for the plus sign, incoming for the minus sign. The surfaces of constant phase are spheres, as are the surfaces of constant amplitude ($A/r = \text{const.}$).

Plane and spherical waves do not, of course, exhaust all the possibilities, but they are the simplest waves, and hence the starting point for understanding more complicated waves. Both

of these types of wave considered here are perfectly coherent in the following sense. The coherence time of a single wave is the greatest time interval for which the phase difference is more or less constant; the coherence length along the direction of propagation is the greatest space interval for which the phase difference is more or less constant. The coherence length is the product of the phase speed of the wave and the coherence time. Thus Eqs. (3.53) and (3.59) describe waves with infinite coherence times, which means that such waves are idealizations never realized in nature. Even a laser beam does not have an infinite coherence time. To do so it would have to have been turned on at the beginning of time and shine until the end of time. So by an infinite coherence time is meant a time much greater than the period (inverse frequency) of the wave.

3.4.5 Acoustic Waves

We noted that Eq. (3.52) is the governing equation for acoustic waves. Waves on a string can be looked upon as one-dimensional acoustic waves in that they are governed by this equation in one dimension. And waves on the strings of musical instruments excite three-dimensional acoustic waves in the air surrounding them.

Although our primary interest is (vector) electromagnetic waves, acoustic waves in fluids are scalar waves and hence simpler. For this reason we often draw analogies between acoustic and electromagnetic waves. Although the two are similar, they are also different, most notably in the way they usually are detected, including by humans. Detectors of light, such as our eyes and photomultiplier tubes, are power detectors: the detected signal is the time-averaged power because of the very short period (inverse frequency) of light waves relative to the detector response. But the *instantaneous amplitude* of sound waves is detected by the human ear.

Suppose that two time-harmonic waves of different frequency but equal amplitude are superposed:

$$a \cos \omega_1 t + a \cos \omega_2 t. \quad (3.60)$$

By using the identities

$$\cos^2(x/2) = \frac{1 + \cos x}{2}, \quad \sin^2(x/2) = \frac{1 - \cos x}{2}, \quad (3.61)$$

$$\cos\left(\frac{x \pm y}{2}\right) = \cos \frac{x}{2} \cos \frac{y}{2} \mp \sin \frac{x}{2} \sin \frac{y}{2}, \quad (3.62)$$

we can write Eq. (3.60) as

$$2a \cos\left\{\frac{(\omega_1 - \omega_2)t}{2}\right\} \cos\left\{\frac{(\omega_1 + \omega_2)t}{2}\right\}. \quad (3.63)$$

If the two frequencies are relatively close to each other we may interpret this equation as describing a wave of frequency equal to the average of the two frequencies, called the *carrier frequency*, with amplitude varying (modulated) at a frequency equal to half the difference of the two frequencies. This frequency difference is called the *beat frequency*. Two harmonic sound waves of different frequency when superposed give rise to the phenomenon of *beats*. And the human ear often can hear these beats.

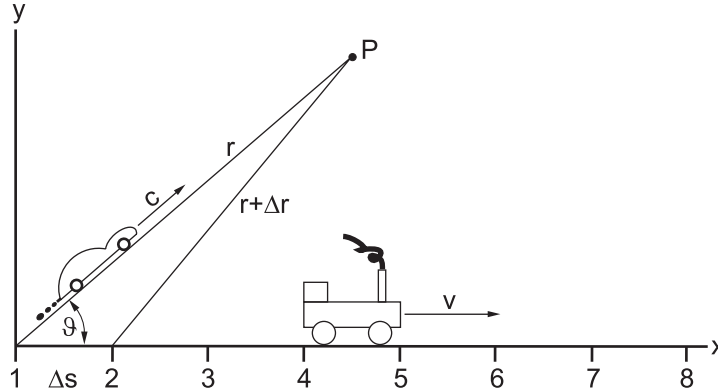


Figure 3.4: Letters are thrown from a train with constant speed v to the drivers of cars waiting at equally spaced stations. These cars carry the letters at constant speed c to a distant point P . In general, the frequency at which letters are sent is not the same as the frequency at which they are received.

But beats are not the sole province of acoustics. In radio engineering producing beats is called *heterodyning*, from the Greek *hetero* meaning different and *dynamis* meaning power. At optical frequencies, mixing of waves of two different frequencies to produce a measurable beat frequency is called *optical beating*, *optical mixing*, or *optical heterodyning*. This has atmospheric applications, most notably the measurement of wind velocities by exploiting the *Doppler effect*, which we turn to next.

3.4.6 The Doppler Effect

The Doppler effect, briefly stated, is the difference between the frequency of transmission of something – call it a signal – and the frequency of its reception because of relative motion between the source and receiver of the signal. There is nothing inherent in the Doppler effect that requires the signal to be transmitted by an acoustic, electromagnetic, or any other kind of wave. To show this, consider the following example inspired by Thomas P. Gill's excellent book, *The Doppler Effect*: a train carrying letters is moving with constant speed v along a straight section of track (Fig. 3.4). Stations are spaced at regular intervals Δs . At each station a car is sitting with its engine running. As the train passes a station, a letter is thrown to the driver of the car, who then drives at constant speed c to a distant point P where the letter is delivered. The time between sending letters is constant, as is the time between receiving them. But the sender of the letters and their recipient at P disagree about these time intervals.

The time t'_1 at which the first letter is received is greater than the time t_1 at which it was sent because of the travel time over the distance r between the first station and P :

$$t'_1 = t_1 + \frac{r}{c}. \quad (3.64)$$

And the time t'_2 at which the second letter is received is greater than the time t_2 at which it was sent:

$$t'_2 = t_2 + \frac{r + \Delta r}{c}, \quad (3.65)$$

where $r + \Delta r$ is the distance between the second station and P . The time interval between sending the two letters is Δt , that between receiving them is $\Delta t'$:

$$\Delta t = t_2 - t_1, \quad \Delta t' = t'_2 - t'_1. \quad (3.66)$$

Combine Eqs. (3.64)–(3.66) to obtain

$$\Delta t' = \Delta t + \frac{\Delta r}{c}. \quad (3.67)$$

This equation is the Doppler effect in a nutshell: $\Delta t' \neq \Delta t$ for a finite speed of transmission.

Let the origin of the xy -coordinate system be at the first station. The coordinates of P are $(r \cos \vartheta, r \sin \vartheta)$. We can find Δr from

$$r + \Delta r = \sqrt{(r \cos \vartheta - \Delta s)^2 + (r \sin \vartheta)^2} = r \sqrt{1 - \frac{2\Delta s \cos \vartheta}{r} + \left(\frac{\Delta s}{r}\right)^2}. \quad (3.68)$$

If the separation between stations is small compared with r , we can expand the right side of Eq. (3.68) in a power series and retain only the first two terms:

$$\Delta r \approx -\Delta s \cos \vartheta + \frac{1}{2} \frac{(\Delta s)^2}{r}. \quad (3.69)$$

The separation between stations is

$$\Delta s = v \Delta t. \quad (3.70)$$

Equations (3.67), (3.69), and (3.70) yield

$$\frac{\Delta t'}{\Delta t} = 1 - \frac{v}{c} \cos \vartheta + \frac{1}{2} \frac{v}{c} \frac{\Delta s}{r}. \quad (3.71)$$

The limit of this quotient as Δs approaches zero (zero separation between stations) is

$$\lim_{\Delta s \rightarrow 0} \frac{\Delta t'}{\Delta t} = 1 - \frac{v \cos \vartheta}{c}. \quad (3.72)$$

The frequency of transmission of letters is $\nu = 1/\Delta t$; their frequency of reception is $\nu' = 1/\Delta t'$. Thus the ratio of frequencies (in the limit of zero separation between stations) is

$$\frac{\nu'}{\nu} = \left(1 - \frac{v \cos \vartheta}{c}\right)^{-1}. \quad (3.73)$$

If $v \ll c$ (slow train, fast car), this ratio is approximately

$$\frac{\nu'}{\nu} \approx 1 + \frac{v \cos \vartheta}{c}. \quad (3.74)$$

Note that $\cos \vartheta$ can be positive or negative. When it is positive the transmitter (source) is approaching the receiver (observer) and the frequency of reception ν' is *greater* than the frequency of transmission ν . This frequency shift is sometimes called a *blue shift* because the frequency of visible light would be shifted upward (toward the blue) when the distance between the transmitter and receiver is decreasing. When $\cos \vartheta$ is negative, the transmitter is receding from the receiver and the frequency of reception is *less* than the frequency of transmission. This frequency shift is sometimes called a *red shift* because the frequency of visible light would be shifted downward (toward the red). But a blue shift has nothing to do with the blue sky, nor a red shift with red sunsets. The terms blue shift and red shift are used even for signals propagated by waves that do not evoke sensations of color (e.g., acoustic waves).

For electromagnetic waves c in Eq. (3.74) is the speed of light. This equation is not relativistically correct, but the correction is small if $v \ll c$, which it is in all the atmospheric applications we know about. According to Eq. (3.74) there is no frequency shift of electromagnetic radiation when $\cos \vartheta = 0$. But according to relativity theory there is a shift (*transverse* Doppler shift), although appreciably smaller than the *longitudinal* Doppler shift.

We have to be careful when applying Eq. (3.74) to acoustic waves because, unlike electromagnetic waves, they propagate relative to a medium (e.g., air) which itself can be moving relative to transmitter and receiver.

For electromagnetic waves the relative frequency shift from Eq. (3.74) is

$$\frac{\nu' - \nu}{\nu} \approx \frac{v \cos \vartheta}{c}, \quad (3.75)$$

where $v \cos \vartheta$ is the component of the velocity of the receiver relative to the transmitter along the line joining them, positive if the two are approaching, negative if they are receding.

Now consider an air molecule illuminated by a monochromatic beam. The frequency of the beam in a coordinate system stationary with respect to the source is ν . But the molecule is, in general, moving relative to the source. As far as the molecule is concerned it is excited by, and hence scatters, radiation with frequency ν' given by

$$\nu' \approx \nu \left(1 + \frac{v_x}{c} \right), \quad (3.76)$$

where for conciseness we write $v_x = v \cos \vartheta$. But at the source an observer detects scattered radiation of frequency

$$\nu'' \approx \nu' \left(1 + \frac{v_x}{c} \right) \approx \nu \left(1 + \frac{v_x}{c} \right)^2 \approx \nu \left(1 + \frac{2v_x}{c} \right). \quad (3.77)$$

Molecular velocity components in any direction are distributed according to the Maxwell–Boltzmann distribution

$$p(v_x) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{mv_x^2}{2k_B T}\right), \quad (3.78)$$

a member of the same family as the Maxwell–Boltzmann distribution for molecular kinetic energies (Sec. 1.2); m is the molecular mass, k_B Boltzmann's constant, and T absolute temperature. As with all continuous probability distributions the integral of p over any interval is

the probability of a molecule having a velocity component in that interval. From Eq. (3.77) we obtain the *relative* frequency shift

$$\delta = \frac{\nu'' - \nu}{\nu} \approx \frac{2v_x}{c}. \quad (3.79)$$

Knowing the distribution function for v_x we can find that for δ :

$$p(\delta) = p(v_x) \frac{dv_x}{d\delta} = \sqrt{\frac{mc^2}{8\pi k_B T}} \exp\left(-\frac{mc^2 \delta^2}{8k_B T}\right). \quad (3.80)$$

This probability distribution is peaked at and symmetric about $\delta = 0$. Values of δ for which the probability is one-half the peak are obtained from Eq. (3.80):

$$\delta = \pm 2\sqrt{\ln 2} \frac{v_m}{c}, \quad (3.81)$$

where the *most probable molecular speed* is

$$v_m = \sqrt{\frac{2k_B T}{m}}. \quad (3.82)$$

We define the width Δ of the probability distribution for scattered frequencies (observed at the transmitter) as the range of relative frequency shifts for which the probability is at least one-half the peak:

$$\Delta = 4\sqrt{\ln 2} \frac{v_m}{c}. \quad (3.83)$$

At normal terrestrial temperatures the most probable speed of air molecules is about 10^3 m s^{-1} whereas c is close to $3 \times 10^8 \text{ m s}^{-1}$, and hence the relative Doppler shift, upward and downward, upon scattering of monochromatic radiation by moving molecules is of order 10^{-5} .

Doppler radar and lidar are used for measuring wind velocities (strictly, wind velocity components along the line from the transmitter-receiver to the point of observation). For example, particles carried by the wind scatter radiation Doppler shifted relative to the source. Wind speeds are of order 10 m s^{-1} , so the relative Doppler shift is around 10^{-7} . This is not large but it can be measured. Indeed, frequency is the physical quantity that can be measured most precisely, about 1 part in 10^9 .

Although we cannot see the consequences of the Doppler shift for light interacting with moving matter, we can hear the consequences of the Doppler shift of sound waves. The speed of sound in air is around 300 m s^{-1} , and hence a *pure tone* from an object moving relative to an observer at 30 m s^{-1} will be Doppler shifted by about 10%, which can be detected by a human observer who is not tone deaf.

3.4.7 Interference of Waves with Different Directions

We showed that waves of different frequency do not interfere in the sense that the time-averaged power transmitted by them is the sum of the powers transmitted by each wave separately. What about waves with the same frequency but propagating in different directions?

You sometimes encounter assertions that waves propagating in different directions cannot or do not interfere. There are (at least) two interpretations of this assertion. One is unobjectionable. Suppose two beams are crossed. Outside the region of overlap each beam is exactly as it would be if it alone were propagating. Thus in this sense the two beams may be said to not interfere. This property of waves is not shared by two streams of automobiles. If they cross each other, collisions irreversibly change the automobiles. But this use of interfere is more akin to its meaning of bother, disturb, or molest: “Please don’t interfere with my plans.” In optics, interfere has a more restricted meaning, namely, waves interfere if the time-average power transmitted by their superposition is *not* the sum of the powers transmitted by each separately. Let us consider in this sense interference of two plane waves with the same frequency and amplitude but different directions. Superposition of two such waves yields

$$a \exp(i\mathbf{k}_1 \cdot \mathbf{x} - i\omega t) + a \exp(i\mathbf{k}_2 \cdot \mathbf{x} - i\omega t + i\Phi). \quad (3.84)$$

Given the phase difference

$$\Delta\varphi + \Phi = (\mathbf{k}_2 - \mathbf{k}_1) \cdot \mathbf{x} + \Phi \quad (3.85)$$

between these two waves, where Φ is a constant, what are we to make of assertions that waves with the same frequency propagating in different directions do not interfere?

To answer this choose $\mathbf{k}_1 = k\mathbf{e}_z$ and $\mathbf{k}_2 = \mathbf{k}_1 + \Delta\mathbf{k}$. These two directions determine a plane, which we take to be the yz -plane. Because the magnitude of both wavevectors is the same we have

$$\Delta\mathbf{k} = k \sin \vartheta \mathbf{e}_y + k(\cos \vartheta - 1)\mathbf{e}_z, \quad (3.86)$$

where ϑ is the angle between the two wavevectors. The phase shift $\Delta\varphi$ is therefore

$$\Delta\varphi = ky \sin \vartheta + kz(\cos \vartheta - 1). \quad (3.87)$$

Suppose that the detector is a disc centered at the origin and lying in the xy -plane. The power per unit area at each point of the detector is therefore [see Eq. (3.42)]

$$2a^2\{1 + \cos(\Delta\varphi + \Phi)\} = 2a^2\{1 + \cos \Delta\varphi \cos \Phi - \sin \Delta\varphi \sin \Phi\}, \quad (3.88)$$

where

$$\Delta\varphi = 2\pi y \sin \vartheta / \lambda, \quad (3.89)$$

and y lies between $-R$ and R , the radius of the detector. To simplify further, assume that $\vartheta \ll 1$ (i.e., the waves are almost in the same direction), which allows us to approximate $\sin \vartheta$ as ϑ . With this approximation Eq. (3.88) becomes

$$2a^2\{1 + \cos(2\pi y \vartheta / \lambda) \cos \Phi - \sin(2\pi y \vartheta / \lambda) \sin \Phi\}. \quad (3.90)$$

The cosine term oscillates between positive and negative values, its argument lying between $-2\pi R \vartheta / \lambda$ and $2\pi R \vartheta / \lambda$. The integral of the sine term over the circular disc vanishes. If the radius of the detector is such that the range of $2\pi y \vartheta / \lambda$ is much greater than 2π , that is, if

$$R \gg \frac{\lambda}{2\vartheta}, \quad (3.91)$$

then if Eq. (3.90) is integrated over the detector, all the positive oscillations of the cosine almost cancel all the negative oscillations, leaving the result that the total power received by the detector is $2a^2$, the sum of the powers of the two waves acting separately.

The condition Eq. (3.91) is likely to be satisfied at visible wavelengths. Suppose, for example, that the waves diverge by one degree. When superposed they will not yield an interference pattern if the radius of the detector is appreciably greater than $10\text{ }\mu\text{m}$. But we can think of two uncontrived examples in which Eq. (3.91) is *not* satisfied. As we explain in the following subsection, the angular pattern of scattering by a particle is an interference pattern. A particle may be looked upon as a coherent array of dipoles excited to radiate waves by an incident beam. What is measured in a particular direction by a detector is the superposition of all these waves. Although they propagate in (slightly) different directions they still interfere. In typical laboratory experiments the detector may be at a distance of 0.5 m from a particle in the sample. Suppose that the linear size of a particle is $10\text{ }\mu\text{m}$ or less. The angular spread of the waves from all the dipoles making up the particle is therefore around 10^{-5} rad . This yields a value for $\lambda/2\vartheta$ of at least 1 cm , which is comparable with the linear dimensions of the detector. That is, for a detector at 0.5 m to resolve angular differences of about a degree the detector must be around 1 cm in diameter.

Now consider two identical particles. One may be looked upon as a detector of radiation from the other (and vice versa, of course). Suppose the particles are cloud droplets, which have typical dimensions of $10\text{ }\mu\text{m}$. The average separation of cloud droplets is around $2000\text{ }\mu\text{m}$, which corresponds to an angular spread of dipolar waves from one particle at the other particle of about $5 \times 10^{-3}\text{ rad}$. At visible wavelengths $\lambda/2\vartheta$ is therefore around $50\text{ }\mu\text{m}$. But the dimensions of the detector (the other particle) are smaller than this. Thus as far as this particle in its role as a detector is concerned, it receives waves from different directions that do indeed interfere. Waves from dipolar oscillators making up particles are not plane waves, closer to being spherical waves, which at sufficiently large distances can be considered to be locally planar (i.e., plane over a limited area). Nor are they all of equal amplitude. But these departures from the ideal do not change our general conclusion that waves in different directions can interfere.

3.4.8 Phase Shift on Scattering

Consider two identical dipoles (Fig. 3.5) illuminated by a monochromatic (scalar) plane wave, or by a source with a lateral coherence length much greater than the separation between them, with wavenumber $k = 2\pi/\lambda$. Assume that they are excited mostly by the incident wave, mutual excitation being negligible. Because the two dipoles are excited by the same incident wave, the waves scattered by them bear a fixed phase difference that depends on the separation between them, the wavelength of the illumination, and the direction of observation. Denote by \mathbf{e}_i the direction of the incident wave, by \mathbf{e}_{s1} and \mathbf{e}_{s2} the directions of the scattered waves (directions toward the point of observation O), and by \mathbf{r}_{12} the position of dipole 2 relative to dipole 1. Although the scattered waves are more or less spherical waves, we take the point of observation to be sufficiently far away that they can be considered to be locally plane waves. We consider scalar waves even though light is a vector wave because we are interested only in the phase difference. If the complex amplitude of the incident wave at 1 is a , at 2 it is $a \exp(ik\mathbf{e}_i \cdot \mathbf{r}_{12})$. Because the amplitude of each scattered wave is proportional to the

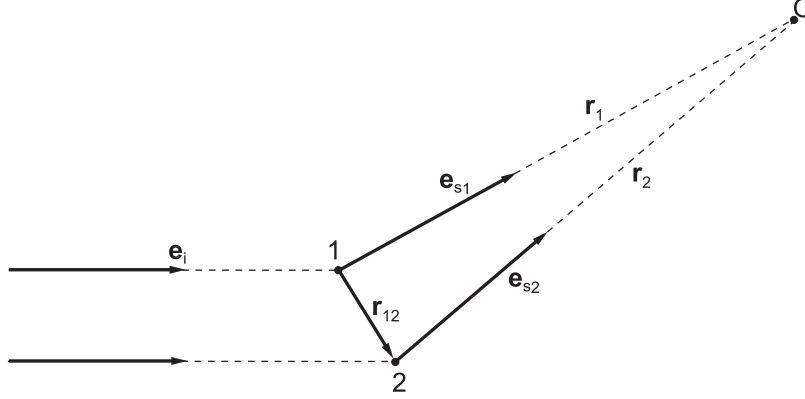


Figure 3.5: At O the total scattered field is the superposition of fields scattered by the two dipoles a fixed distance r_{12} apart. The phase difference between these two fields depends on the angle between the incident and scattered directions and the distance between the dipoles relative to the wavelength of the illumination.

amplitude of the wave that excites it, the two scattered waves at O are

$$\psi_1 = a \exp(ik\mathbf{e}_{s1} \cdot \mathbf{r}_1), \quad \psi_2 = a \exp(ik\mathbf{e}_i \cdot \mathbf{r}_{12} + ik\mathbf{e}_{s2} \cdot \mathbf{r}_1 - ik\mathbf{e}_{s2} \cdot \mathbf{r}_{12}), \quad (3.92)$$

where

$$\mathbf{r}_2 = \mathbf{r}_1 - \mathbf{r}_{12}. \quad (3.93)$$

Any factors common to the two waves are omitted. As the distance to the observation point O increases indefinitely, $\mathbf{e}_{s2} \rightarrow \mathbf{e}_{s1} = \mathbf{e}_s$, and the phase difference between the two waves approaches

$$\Delta\varphi = k(\mathbf{e}_i - \mathbf{e}_s) \cdot \mathbf{r}_{12} = \frac{2\pi}{\lambda}(\mathbf{e}_i - \mathbf{e}_s) \cdot \mathbf{r}_{12}. \quad (3.94)$$

From this simple equation flows an amazing amount of physical understanding.

The *forward direction* is the direction of the incident wave. According to Eq. (3.94), scattering in the forward direction ($\mathbf{e}_s = \mathbf{e}_i$) is always in phase ($\Delta\varphi = 0$) regardless of the separation of the two dipoles and the wavelength. This is the only scattering direction for which this is true, and hence the forward direction is singular, a point to which we return.

Any particle is a coherent array of N dipoles. Again, we assume that they are excited mostly by the incident wave. If the linear extent of this array is small compared with the wavelength, all the separate scattered waves are approximately in phase for all directions [see Eq. (3.42) with $\Delta\varphi = 0$ and $a_1 = a_2$], and hence the power scattered in any direction by N scatterers is N^2 times the power scattered by one. If the total volume of the particle is v and there are n dipoles per unit volume, the total number of dipoles is proportional to v . We therefore predict that the total power scattered by a particle small compared with the wavelength is proportional to v^2 . As we show in Section 3.5, this turns out to be correct.

Interactions between dipoles are not negligible unless separated by sufficiently large distances. Although this condition is not satisfied by a small particle, scattering still increases as the square of its volume even when interactions are accounted for (by interactions is meant that the wave from each dipole appreciably excites its neighbors). The reason for this is that if a particle is sufficiently small, appreciable phase differences are not possible even with interactions. Hence the array of dipoles (particle) is approximately coherent and in-phase even though the dipoles excite each other.

Let us continue in this vein and see what general results we can obtain by physical reasoning based on Eq. (3.94). As the overall size of the particle increases, the phase differences between the waves scattered by its constituent dipoles also increase. When the particle is sufficiently small, all these waves are approximately in phase (constructive interference) for all scattering directions. But as the particle size increases, phase differences of order π or greater become possible for some directions. In particular, the phase difference between two or more dipolar waves can be such that in a particular direction they interfere destructively. Thus we conclude that although scattering by a particle increases as the square of its volume when the particle is small compared with the wavelength, this rapid increase cannot be sustained indefinitely. As particle size increases, scattering increases, but more slowly than volume squared. And again, this expectation is supported by detailed calculations (see Sec. 3.5.1). We can understand them, at least qualitatively, by simple phase difference arguments.

We show in Section 3.2 that scattering by a dipole (subject to restrictions) is approximately inversely proportional to the fourth power of the wavelength of the incident illumination. But this does not necessarily imply that scattering by a coherent array of such dipoles (a particle) also follows this wavelength dependence. As evidenced by Eq. (3.94), an additional wavelength dependence creeps in by way of the phase difference. Indeed, for sufficiently large particles, the net effect of the wavelength dependence of scattering by the individual dipoles nearly cancels the wavelength dependence originating from phase differences. In the following section we give a simple example to elaborate on this point.

Now let us turn to the directional dependence of scattering. Suppose that our two dipoles individually scatter the same in all directions. This does not then imply that scattering by the two together is the same in all directions. According to Eq. (3.94) the phase difference between the two scattered waves depends on the scattering direction. We can make this clearer by considering the special example of two dipoles on a line parallel to the incident wave ($\mathbf{r}_{12} = r\mathbf{e}_i$), in which instance the phase shift (again, interactions are neglected) is

$$\Delta\varphi = \frac{2\pi r}{\lambda}(1 - \cos\vartheta), \quad (3.95)$$

where ϑ is the *scattering angle* (angle between incident and scattered waves). The power scattered in any direction by the two dipoles is determined by $\cos\Delta\varphi$. The quantity $1 - \cos\vartheta$ lies between 0 (forward direction) and 2 (backward direction), and hence the phase difference lies between 0 (forward) and $4\pi r/\lambda$ (backward). When the (backward) phase difference is an odd multiple of π , interference is destructive; when the phase difference is an even multiple of π , interference is constructive; and, of course, everything between these two extremes is possible. As r/λ increases so does the number of oscillations in the scattering diagram (scattering as a function of angle).

As a particle increases in size so does the average distance between its elements. We therefore expect, on the basis of the behavior of two dipoles, the scattering diagram for a particle to exhibit more maxima (minima) the greater its size relative to the wavelength. This expectation is borne out by detailed calculations (Sec. 3.5.1) as well as by measurements.

Because scattering is in-phase in the forward direction regardless of the wavelength and the separation between dipoles, we expect scattering in this direction to increase more rapidly with size than scattering in any other direction. Again, this expectation is borne out by calculations (Sec. 3.5.1) and measurements. A general result is that the larger the particle, the more that scattering by it is peaked in the forward direction.

The rate of change of the phase difference with respect to r

$$\frac{\partial \Delta\varphi}{\partial r} = \frac{2\pi}{\lambda}(1 - \cos \vartheta) \quad (3.96)$$

ranges from 0 in the forward direction to $4\pi/\lambda$ in the backward direction. The implication of this is that if we have a fixed number of dipoles (fixed particle volume) and we move them around (change the particle shape), scattering is least affected in the forward direction and most affected in the backward direction. Again, this is supported by calculations and measurements.

Much of the basic physics of scattering by particles is embodied in the simple phase difference given in Eq. (3.94). The rest is details, some of which we give in Section 3.5, but first we discuss scattering by air and by water.

3.4.9 Scattering by Air and Liquid Water Molecules

Air molecules near sea level are separated on average by about 3 nm (~ 10 molecular diameters). That is, the average volume allocated to each of N molecules in a volume V is V/N , the cube root of which is defined as the average separation. The separation r_{ij} between the i^{th} and j^{th} molecules, however, is distributed statistically from some minimum (approximately a molecular diameter) to some maximum (approximately the cube root of V). Because the lateral coherence length of sunlight is around $50 \mu\text{m}$, we have to add the waves, taking due account of phases, scattered by all the air molecules (about 10^9) in a volume with approximately this linear dimension, which is about 100 times the wavelengths of visible light.

Scattering by Spatially Uncorrelated Molecules

Suppose that N time-harmonic waves, all with the same amplitude a but different phases, are superposed. For sake of argument we take them to be plane waves, and omit the common factor $\exp(ikx - i\omega t)$. The sum of these waves is

$$\Psi = \sum_{j=1}^N a \exp(i\varphi_j). \quad (3.97)$$

Because these are time-harmonic waves, the time-averaged power transmitted by their sum is proportional to

$$\Psi\Psi^* = |\Psi|^2 = a^2 \sum_{j=1}^N \sum_{m=1}^N \exp\{i(\varphi_j - \varphi_m)\} = a^2 N + a^2 \sum_{j=1}^N \sum_{m \neq j}^N \exp\{i(\varphi_j - \varphi_m)\}. \quad (3.98)$$

The sum in the rightmost side of this equation is the addition of phasors (vectors in the complex plane) of unit length but with different polar angles (phase differences $\Delta\varphi_{jk} = \varphi_j - \varphi_k$). We take Eq. (3.98) to be the total light scattered by the N air molecules in a volume equal to the coherence length cubed. The phase differences are given by Eq. (3.94), and hence Eq. (3.98) becomes

$$|\Psi|^2 = a^2 N + a^2 \sum_{j=1}^N \sum_{m \neq j}^N \exp\{ik(\mathbf{e}_i - \mathbf{e}_s) \cdot \mathbf{r}_{jm}\}. \quad (3.99)$$

Air molecules move and their positions are almost completely uncorrelated. That is, a molecule can occupy almost any position in space regardless of the positions of other molecules because the volume accessible to a molecule is about 1000 times its volume. This in turn implies that all phase differences are equally probable, and hence the sum in Eq. (3.99) is approximately zero, which is readily evident if you draw a great many phasors with random directions emanating from a point, similar to the spokes in a bicycle wheel (laced carelessly). Because the positions of air molecules are uncorrelated, scattering by N (in a small volume) is N times scattering by one. As far as scattering by air molecules is concerned, it is *as if* phase does not exist. This is true for all scattering directions except the forward direction ($\mathbf{e}_s = \mathbf{e}_i$), for which scattering is in-phase regardless of the molecular separation relative to the wavelength. In fact, this in-phase scattering is the source of refraction (see Sec. 8.3 on atmospheric refraction and mirages). Historically, refraction has been treated separately from scattering, as if the two had nothing to do with each other. This is not true. Refraction is scattering; it can be looked upon as the coherent part of scattering. Scattered light not associated with refracted light is the incoherent part.

In 1899 Lord Rayleigh attributed the blue of the sky (see Sec. 8.1) to scattering by air molecules. Underlying his theoretical expression for the amount of light scattered by N air molecules was the assumption that the phases of the separate scattered waves are “entirely at random”, and hence scattering by N air molecules is N times scattering by one. But he also recognized that “When the volume occupied by the molecules is no longer very small compared with the whole volume, the fact that two molecules cannot occupy the same space detracts from the random character of the distribution.”

Scattering by air molecules is almost, but not quite, the same in all directions (isotropic), the reason for which is best left for the chapter on polarization (see Sec. 7.3; also Prob. 7.18). Angular scattering by air is also almost, but not quite, the same as that by a small (compared with the wavelength) sphere (see Fig. 3.13), small differences arising because the dominant air molecules are not spherically symmetric.

The Blue of the Sky: Scattering by Fluctuations or by Molecules?

You sometimes encounter the assertion that the blue of the sky (scattering by air molecules) is “really” a consequence of “scattering by fluctuations.” This is piffle, reflecting ignorance of physics and its history, tantamount to denying the existence of molecules. The origins of this piffle go back almost 100 years to the work of Smoluchowski (1908) and Einstein (1910) who developed theories of scattering by dense media (e.g., liquids) by considering matter to be continuous but with spatially varying properties to which they applied thermodynamic arguments. Thermodynamics does not *explicitly* invoke molecules and so, of course, neither do the theories of Smoluchowski and Einstein, which, not surprisingly, contain thermodynamic variables such as temperature and isothermal compressibility. But this does not mean that Smoluchowski and Einstein believed that molecules are not the agents responsible for scattering. Einstein, in particular, recognized that his theory circumvented the difficulties of a molecular theory of scattering in fluids. He noted after his labors that “It is remarkable that our theory does not make *direct* use of the assumption of a discrete distribution of matter.” But the fluctuation theories of Smoluchowski and Einstein have been distorted over the years into the fatuous notion that fluctuations, not molecules, do the scattering.

For an ideal gas, the theories of Smoluchowski and Einstein yield Rayleigh’s result (which Einstein acknowledged) that scattering by N molecules is N times scattering by one, but their theories go further. In particular, they account for *critical opalescence*. At the critical point, the distinction between gas and liquid disappears and scattering greatly increases (according to theory it becomes infinite) as a fluid teeters between the liquid and gas phases. We give Einstein’s scattering formula in Problem 5.16 and include a few problems that require this formula.

Many years later in 1945 Bruno Zimm tackled the problem of scattering in dense media by explicitly considering scattering by molecules. Zimm’s own words demonstrate that he understood this scattering whereas the it-is-really-scattering-by-fluctuations folks do not: “the difficulty [of calculating interference between waves scattered by different molecules] was elegantly circumvented by Smoluchowski and Einstein, who considered the liquid as a continuous medium troubled by small statistical fluctuations in density. The extent of these fluctuations could be calculated from the macroscopic compressibility of the medium, and the intensity of the scattered light was obtained without discussing the individual molecules at all.”

What Zimm did, in essence, was evaluate the sum in Eq. (3.99), approximating it by an integral and accounting for the correlations between molecular separations \mathbf{r}_{ij} by using results from statistical mechanics. The first term in Eq. (3.99) is scattering by N isolated molecules, the second term a consequence of interference of waves from molecules correlated in position. Zimm’s molecular theory of scattering reproduces the results of Einstein and Smoluchowski but goes a step further in that it does not yield infinite scattering at the critical point.

Now let us turn to an ordinary glass of water, absolutely free of contamination (setting aside the difficulty of preparing such water). Molecules in liquid water are separated by distances comparable with their diameter, and hence one molecule cannot move without pushing others aside. Because of this correlation in position, scattering by N water molecules is not simply N times scattering by one. Indeed, scattering per molecule in the liquid phase is appreciably less than in the gas phase. When clean water is illuminated by a monodirectional

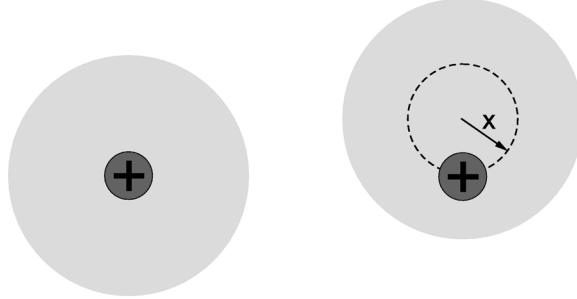


Figure 3.6: Simple model of an atom as a stationary positive nucleus (black sphere) surrounded by an equal but oppositely charged continuous (gray) electronic cloud. When the cloud is displaced a distance x , it experiences an electrostatic restoring force.

source of light, what is observed is described to good approximation by the laws of specular reflection and refraction. Although textbooks rarely say so, these are laws of scattering by coherent arrays. But there is always a residual incoherent component to the total scattering. When water is illuminated, most (but not all) of what is observed is accounted for by the laws of specular reflection and refraction. In particular, the reflected and refracted rays all lie in the plane determined by the normal to the surface of the water and the direction of the incident wave. Yet if we look in directions outside this plane we can observe scattered light, weak but measurable, and not necessarily the result of junk in the water. Water is intrinsically junky in that it is made up of discrete pieces (molecules), highly correlated pieces to be sure but not perfectly correlated.

Frequency Dependence of Scattering by Air Molecules

Light from the sky is a consequence of scattering by molecules essentially uncorrelated in position. The spectrum of this light is given approximately by Eq. (3.9), Rayleigh's scattering law. Air molecules are certainly small enough compared with the wavelengths of visible light that this condition for the validity of the law is satisfied. But there is another condition, evident from Eq. (3.8): the frequencies for which Eq. (3.9) is a good approximation must be much less than a resonant frequency of the scatterer. What is this frequency for an air molecule?

To answer this question we resorted to a crude model of an atom: a continuous (negative) charge distribution (electron cloud) surrounding a positively charged point nucleus (see Fig. 3.6). When the center of the electron cloud coincides with the nucleus, there is no net force, but when the cloud is displaced it experiences a restoring force. Let q be the charge on the nucleus. The charge density of the electron cloud is therefore $-3q/4\pi R^3$, where R is the atomic radius. When the cloud is displaced a distance x , it experiences an attractive force that pulls it back toward the center. This force is that between two point charges, the nucleus and one at the center with charge equal to the total (negative) charge within the sphere of radius x , which is $-qx^3/R^3$. From Coulomb's law we therefore have for the restoring force

$$F = \frac{q}{4\pi\epsilon_0 x^2} \left(\frac{-qx^3}{R^3} \right) = -\frac{q^2 x}{4\pi\epsilon_0 R^3} = Kx, \quad (3.100)$$

where ε_0 (the permittivity of free space) is a constant (8.85×10^{-12} in SI units). This equation should look familiar: it is the restoring force for a linear harmonic oscillator (Sec. 2.6). Thus the resonant frequency follows immediately from Eq. (3.76):

$$\omega_0 = \sqrt{\frac{K}{m}} = \sqrt{\frac{q^2}{4\pi\varepsilon_0 m R^3}}, \quad (3.101)$$

where we take m to be the mass of the electron cloud. If we take $R = 0.15$ nm for the atomic radius (strictly, this is the radius of molecular nitrogen), we obtain a resonant frequency that corresponds to a wavelength of about $0.1 \mu\text{m}$, well into the ultraviolet.

Eq. (3.101) is an example of a *plasma frequency*. Plasma frequencies pop up all over the place, including in the ionosphere, for which plasma frequencies correspond to wavelengths of tens of meters. The plasma frequency is a cutoff frequency: below it a medium is reflecting, above it transmitting. This is why reflection of radio waves (for frequencies below the plasma frequency) by the ionosphere can be put to use for communication on Earth. To communicate beyond Earth would require waves above the plasma frequency. The late cosmologist, Fred Hoyle, wrote *The Black Cloud*, considered to be a classic of science fiction. Earth is menaced by an intelligent interstellar cloud that communicates with scientists on Earth by adjusting its plasma frequency.

3.5 Scattering by Particles

Although the form of Eq. (3.10) is the same regardless of the illuminated object, we use different symbols for molecular and particle scattering cross sections just as we did for absorption cross sections in order to emphasize that molecules and particles are different. Most of what we said in Section 2.8 about particle absorption cross sections also applies to particle scattering cross sections. But unlike absorption, scattering can be apportioned into directions. The *differential scattering cross section*, often written as $dC_{\text{sca}}/d\Omega$, is the contribution to the total scattering cross section C_{sca} from scattering into a unit solid angle (Sec. 4.1.1) in each direction, and hence

$$C_{\text{sca}} = \int_{4\pi} \frac{dC_{\text{sca}}}{d\Omega} d\Omega, \quad (3.102)$$

where integration is over all directions (4π steradians). The differential scattering cross section is *not* the derivative of C_{sca} with respect to the variable Ω , the notation being chosen merely to aid our memories.

The *scattering coefficient* β of a suspension of N identical particles per unit volume is defined similarly to its absorption coefficient:

$$\beta = NC_{\text{sca}}, \quad (3.103)$$

an expression readily generalized to a suspension of non-identical particles by adding their individual contributions to the total scattering coefficient. The sum of the absorption and scattering cross sections is called the *extinction cross section*:

$$C_{\text{ext}} = C_{\text{abs}} + C_{\text{sca}}, \quad (3.104)$$

and the sum of absorption and scattering coefficients is called the *extinction coefficient*. As with absorption cross sections, those for scattering and extinction are sometimes normalized by their geometrical (projected) cross-sectional areas G to yield dimensionless efficiencies or efficiency factors for scattering and extinction:

$$Q_{\text{sca}} = \frac{C_{\text{sca}}}{G}, \quad Q_{\text{ext}} = \frac{C_{\text{ext}}}{G}, \quad (3.105)$$

neither of which has any physical significance (whereas Q_{abs} does) although they are computationally convenient. As we argued about absorption cross sections, volumetric scattering and extinction cross sections (for particles with well-defined volumes) are better measures of the efficiency of a particle at scattering and extinguishing radiation.

For a suspension of N identical particles Eq. (3.103) becomes [see Eq. (2.143)]

$$\beta = \frac{f C_{\text{sca}}}{v}, \quad (3.106)$$

where v is the volume of a single particle and f is the volume fraction of particles. For a suspension of nonidentical particles [see Eqs. (2.144)–(2.145)]

$$\beta = f \left\langle \frac{C_{\text{sca}}}{v} \right\rangle, \quad (3.107)$$

where f is the total volume fraction of all particles and the brackets indicate an ensemble average over all particle types. Underlying Eqs. (3.103) and (3.107) is the assumption that the consequences of interference are negligible (i.e., the suspension is an incoherent array). We know, however, that this cannot be true for arbitrary f . In the limit $f \rightarrow 1$ (if the distribution of size and shapes of particles is such that they can fill all space) the scattering coefficient should plummet (see Sec. 3.4.9 for the differences between scattering per molecule by air and by liquid water). But Eq. (3.107) predicts (incorrectly) a steadily increasing scattering coefficient. The same cautionary statements we made about absorption coefficients of suspensions are applicable to scattering coefficients.

In a transmission measurement the *sum* of absorption and scattering (i.e., extinction) is unavoidably measured. If multiple scattering is negligible a monodirectional beam is attenuated exponentially (Sec. 5.2.3):

$$F = F_0 \exp\{-N \langle C_{\text{ext}} \rangle x\}. \quad (3.108)$$

If we measure the ratio of transmitted to incident irradiances, knowing the path length and the number density of particles, we can, in principle, infer the (average) extinction cross section but we cannot determine the separate contributions of absorption and scattering to extinction. In any scattering medium, absorption is never totally absent; in any absorbing medium scattering is never totally absent. To separate extinction into its components requires either additional kinds of measurements, observations, or guidance from theory.

To get some understanding of how scattering cross sections of particles depend on their size and the wavelength of the illumination, we follow the same path as we did in trying to grasp absorption cross sections (Sec. 2.4). We consider reflection by a (negligibly absorbing)

slab of uniform thickness h , area $A \gg \lambda^2$, and refractive index n , illuminated at normal incidence by a monodirectional, monochromatic beam with irradiance F_0 . For simplicity we take the slab to be in air. The total power reflected by this slab is

$$W_r = F_0 A \tilde{R}, \quad (3.109)$$

where the coherent reflectivity (see Sec. 5.1.1) is

$$\tilde{R} = \frac{2R_\infty(1 - \cos \varphi)}{1 - 2R_\infty \cos \varphi + R_\infty^2}. \quad (3.110)$$

The reflectivity of the infinite (only one boundary) medium is

$$R_\infty = \left| \frac{n-1}{n+1} \right|^2, \quad (3.111)$$

and the phase difference as a consequence of two parallel boundaries a finite distance apart is

$$\varphi = \frac{4\pi n h}{\lambda}. \quad (3.112)$$

We use the term coherent reflectivity to emphasize that, like a particle, a slab is a coherent object.

The reflection cross section is defined as W_r/F_0 , and hence the reflection cross section per unit volume $v = Ah$ is

$$C_{\text{ref}}/v = \frac{\tilde{R}}{h}. \quad (3.113)$$

For a sufficiently thin slab (relative to the wavelength), $\varphi \ll 1$, and Eq. (3.113) is approximately

$$C_{\text{ref}}/v \approx \frac{(n^2 - 1)^2 \pi^2 h}{\lambda^2}. \quad (3.114)$$

This expression resembles what we predicted in Sections 3.1 and 3.4.8 for scattering by a small particle (at wavelengths far from absorption bands), namely volumetric scattering proportional to size (volume) and to wavelength to a power (-4) . A thin slab is not a dipole but can be looked upon as a sheet of dipoles, and hence the different (but similar) dependence of the volumetric reflection cross section on size and wavelength, increasing with increasing size and proportional to a power of the wavelength (-2) . And note that this power is not exactly -2 because n also depends on wavelength. We note in passing that Eq. (3.114) is the source of the incorrect attribution of the blue sky (Sec. 8.1) to reflection by thin plates. Close but not close enough.

We can imagine the slab thickness to increase by piling one sheet of dipoles onto another. At first, the volumetric reflection cross section increases monotonically with each added sheet (Fig. 3.7) as long as all parts of the slab are excited in phase with all other parts. But once the total thickness reaches the point where appreciable ($\sim \pi/2$) phase differences among the

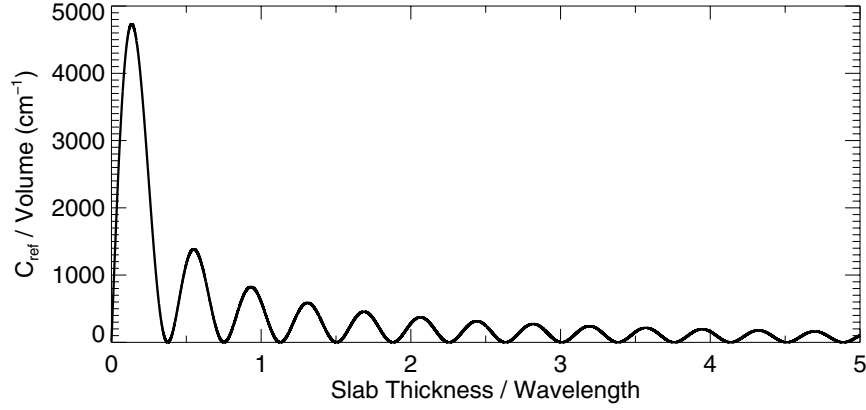


Figure 3.7: Reflection (scattering in the backward direction) cross section per unit volume of a transparent slab (with refractive index 1.33) illuminated at normal incidence as a function of slab thickness relative to the wavelength of the illumination (1 μm).

different parts are possible, the cross section decreases with increasing thickness, then oscillates with decreasing amplitude and decreasing distance between adjacent peaks. Moreover, the wavelength dependence of reflection ceases to resemble that for a dipole sheet. Because of interference, the sum of dipoles is not a dipole, which, in a nutshell, is what makes optics interesting.

Despite appearances to the contrary a plot of C_{ref}/v versus h for fixed frequency is *not* equivalent to a plot versus $1/\lambda$ (frequency) for fixed h . Frequency cannot be increased indefinitely and n remain constant. No such material exists, except a perfect vacuum, which also does not exist. Moreover, frequency cannot be varied arbitrarily without encountering non-negligible absorption (see Figure 3.8). If you pretend that absorption is negligible for more than a decade of wavelengths you probably have slipped into the world of fantasy.

3.5.1 Complex Refractive Index

The theory of scattering and absorption by a homogeneous particle contains a single material parameter, its *complex refractive index*. The simplest way to demonstrate that refractive indices can be written as complex numbers is to consider a plane, scalar wave propagating in the $+x$ direction in an optically homogeneous medium:

$$\psi = a \exp(ikx - i\omega t). \quad (3.115)$$

The phase speed v of this wave is ω/k . If c is the free-space speed of light we can rewrite this as

$$\psi = a \exp\{i\omega(nx/c - t)\}, \quad (3.116)$$

where

$$n = \frac{c}{v} \quad (3.117)$$

is the refractive index of the medium in which the wave propagates. But this wave cannot be physically realistic because it is not attenuated if a is constant. So let's allow the wave to attenuate exponentially by writing its amplitude as

$$a \exp(-n''\omega x/c). \quad (3.118)$$

Now the wave has the form

$$\psi = a \exp\{i\omega(Nx/c - t)\}, \quad (3.119)$$

where the *complex refractive index* is

$$N = n' + in'' \quad (3.120)$$

and we changed the symbol for its real part from n to n' . Alternatively, we can say that the wavenumber in Eq. (3.115) is complex and write it as

$$k = \frac{\omega}{c}N = k' + ik'' \quad (3.121)$$

By now we should be comfortable with complex wavenumbers because we know that solutions to linear equations, such as the wave equations we have considered, can be expressed as complex functions (see Sec. 2.5). Because the time-averaged power transmitted by a wave in a medium with a complex refractive index is (except for a factor)

$$\psi\psi^* = |\psi|^2 = a^2 \exp\{-2n''\omega x/c\} = a^2 \exp\{-4\pi n''/\lambda\}, \quad (3.122)$$

where $\lambda = 2\pi c/\omega$ is the free space wavelength, we immediately obtain the relationship between the imaginary part of the refractive index and the absorption coefficient denoted first by κ in Section 2.1, then changed to κ_b in Section 2.8 to distinguish between bulk (homogeneous) media and suspensions of particles:

$$\kappa_b = \frac{4\pi n''}{\lambda}. \quad (3.123)$$

The symbol α is also used for the bulk absorption coefficient in Eq. (3.123).

The notation and conventions for complex refractive indices sometimes cause confusion and can lead to serious errors. For our time-harmonic convention, both n' and n'' are non-negative. And yet you often encounter complex refractive indices written as

$$N = n' - in'', \quad (3.124)$$

where again both n' and n'' are non-negative. Which is correct, Eq. (3.120) or Eq. (3.124)? Both when used with the appropriate convention. If the convention is $\exp(-i\omega t)$, Eq. (3.120)

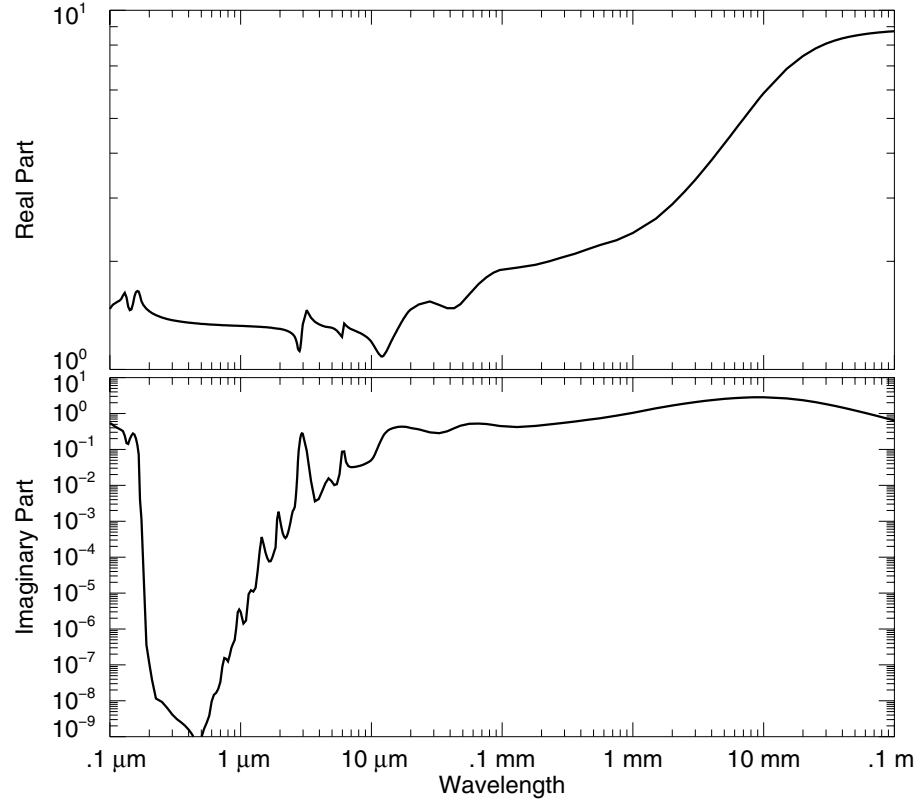


Figure 3.8: Optical constants (complex refractive index) of liquid water taken from the compilation by Querry *et al.* (1991) cited at the end of Chapter 2.

corresponds to waves that decrease in amplitude in the direction of propagation; if the convention is $\exp(i\omega t)$, Eq. (3.124) corresponds to such waves. If we could devise waves that increase in amplitude in the direction of propagation, without an external source feeding energy to them, the energy crisis would immediately evaporate. Conventions are neither right nor wrong. What is wrong is to mix conventions, a sure recipe for accidents (as is driving on the right-hand side of the road in Britain).

The complex refractive index often is written $N = n + ik$, with n and k called *optical constants* (especially by the folks who measure them). But just as the Lord Privy Seal is neither a lord nor a privy nor a seal, optical constants are neither optical (pertain to only visible or near-visible frequencies) nor constant. Indeed, they vary appreciably over frequencies from radio to ultraviolet (Fig. 3.8). Moreover, they are not independent, adjustable parameters that separately can take on any values. The real and imaginary parts are linked. If one varies, so must the other. For example, around $3\ \mu\text{m}$, a large variation in the imaginary part of the refractive index of water (absorption band) is accompanied by a variation in the real part (Fig. 3.8).

Real Refractive Index and Phase Speed

Few concepts are treated worse in textbooks, are encumbered by more myths and demonstrable nonsense, than the refractive index, complex or otherwise. For example, it has been stated countless times that the refractive index (by which is meant the real part) is greater in a denser medium. Yes, it is true that the refractive index of air is less than that of a thousand times denser water at most, maybe all (we don't know for sure), frequencies. But that's about as far as it goes. A delightfully humorous but edifying note about refractive indices and (mass) density was published half a century ago by the late R. Scott Barr in *American Journal of Physics* (which is supposed to be read by physics teachers and definitely must be read by authors of physics textbooks). The caption to Fig. 1 in Barr's note reads "Does index of refraction vary directly with density?" This figure shows refractive index versus density (at a given frequency in the visible) for a few dozen substances. The points are connected to answer the question in the caption: NO. Barr confesses that these points "were not selected entirely impartially", so shows another figure in which refractive index versus density is plotted (for a given temperature and wavelength) for 445 organic materials. As Barr notes, "this figure indicates that if a material has a high density it is more likely to have a high index than a low one, but that is about all that can be said." Actually, the picture is much bleaker than Barr paints. It is no trick at all to find two materials with different densities for which at one frequency the refractive index is greatest for the denser material but at another frequency is least. And now the clincher: there are plenty of dense, solid materials with *lower* refractive indices at some frequencies than that of such diaphanous substances as air. And these materials are not exotic or "anomalous", but close to hand. For example, at a wavenumber of about 530 cm^{-1} , the real part of the complex refractive index of magnesium oxide (made by burning magnesium wire in air) is 0.093. If magnesium oxide is too exotic for your tastes, try glass (silicon dioxide), with a real refractive index of 0.39 at 1250 cm^{-1} . And then there is ordinary table salt (sodium chloride), with a real refractive index of 0.14 at 200 cm^{-1} . Examples abound. We did not have to search high and low for them.

But having mentioned refractive indices near zero, we anticipate howls of outrage from those who have been taught that nothing can travel faster than light. It is indeed true that according to special relativity no *thing* can travel faster than light, but a surface of constant phase is not a thing. It is an imaginary surface moving with the phase speed v . Special relativity also dictates that *signals* cannot be propagated faster than c , but a plane, harmonic wave such as Eq. (3.33) cannot transmit a signal. You likely have an AM-FM radio. AM stands for *amplitude modulation*, FM for *frequency modulation*. Radio stations, despite proclaiming a single broadcast frequency, do not transmit absolutely monochromatic waves. Programs transmitted by such waves would be exceedingly dull: just a steady hum. To transmit a signal (information) requires modulating waves. Consider the simpler example of Morse code. A steady signal would convey no information. To convey information by Morse code requires turning something on for a brief instant, then off, varying the time intervals between off and on and the length of the time on. That is, the steady tone must be modulated. And so must speech in order to be understood. Try carrying on a conversation at a party by steadily grunting at a single frequency. You won't be invited again. Because the phase speed is neither the speed of a thing nor the speed with which signals are propagated, it is not constrained to be less than c , and hence the real part of the refractive index is not constrained to be greater than 1.

Students who stumble upon real refractive indices less than 1 may be placated by being told that the *group velocity* can't be greater than c . Alas, invoking the group velocity only makes matters worse. To understand why, we first need to explain the group velocity. Consider two waves of equal amplitude, propagating together, with wavenumbers $k \pm \Delta k/2$ and frequencies $\omega \pm \Delta\omega/2$, in a medium we take to be nonabsorbing ($N = n$ and k are real) for convenience. The sum of these two waves [after a bit of algebra, see Eq. (3.63)] is

$$\psi = 2a \exp\{ikx - i\omega t\} \cos\{\Delta kx/2 - \Delta\omega t/2\}. \quad (3.125)$$

This describes a plane harmonic wave with phase speed ω/k the amplitude of which is a wave with a much greater wavelength (much smaller wavenumber) and a correspondingly much smaller frequency. The speed of this long wavelength envelope modulating the much shorter wavelength wave is $\Delta\omega/\Delta k$, which leads to the definition of the *group velocity* (speed, really) v_g as the limit

$$\frac{1}{v_g} = \lim_{\Delta\omega \rightarrow 0} \frac{\Delta k}{\Delta\omega} = \frac{dk}{d\omega}. \quad (3.126)$$

From Eqs. (3.117) and (3.121) we have

$$k = \frac{\omega n}{c}, \quad \frac{1}{v} = \frac{k}{\omega} = \frac{n}{c}. \quad (3.127)$$

Combining Eqs. (3.126) and (3.127) yields

$$v_g = \frac{c}{n + \omega(dn/d\omega)}. \quad (3.128)$$

A medium in which n depends on ω is called *dispersive*, but because a truly non-dispersive medium does not exist (except possibly empty space), this term is not very distinctive. But the term dispersion does convey that a group of waves propagating together in a medium will not stick together. To disperse means “to become widely separated.”

If $dn/d\omega = 0$, $v_g = v = c/n$, and a group of waves of different frequency and wavenumber would all have the same phase speed and hence would not disperse. We can rewrite Eq. (3.128) as

$$v_g = \frac{v}{1 + (\omega/n)(dn/d\omega)}. \quad (3.129)$$

The term *normal dispersion* is applied when $dn/d\omega > 0$, *anomalous dispersion* when $dn/d\omega < 0$, but there is nothing “anomalous” about anomalous dispersion. All materials exhibit anomalous dispersion at some frequencies, and when everything is anomalous, nothing is anomalous. In a frequency region of normal dispersion, $v_g < v$; in a region of anomalous dispersion, $v_g > v$. When we carefully examine Eq. (3.128) we arrive at a startling conclusion. Suppose that

$$n + \omega \frac{dn}{d\omega} < 0, \quad (3.130)$$

which can be satisfied only in a region of anomalous dispersion, and that

$$\left| n + \omega \frac{dn}{d\omega} \right| < 1. \quad (3.131)$$

If these two conditions are satisfied the group velocity is negative and its magnitude is greater than c : light breaks the speed limit while traveling the wrong way. This is why we said that invoking the group velocity when faced with phase speeds greater than c only makes matters worse. Are there any real materials that satisfy Eqs. (3.130) and (3.131)? Tons of them, as near as your breakfast table: the group velocity in table salt at a frequency of 200 cm^{-1} is about $-4c$.

Alleged violations of the speed limit c are a consequence of the careless, if not downright sloppy, ways in which the term speed of light is used. There are (at least) four different speeds of light: the phase speed, the group speed, the signal speed, and the speed of energy propagation. All four are different in a dispersive medium, but only one of them cannot be greater than c . As its name implies, the signal speed is the speed with which signals are propagated. Suppose that a source in a dispersive medium is suddenly turned on. After time t a detector a distance d from the source begins to register a signal. Special relativity says that the ratio d/t cannot be greater than c . As with Morse code, signals can be transmitted by light only by turning sources on and off.

Imaginary Refractive Index and Absorption by Particles

Although the real part of the complex refractive index is surrounded by misconceptions, when we turn to its imaginary part we encounter so much nonsense that our heads spin. If you think we exaggerate, consider the following story, which we did not make up (or could have). Several years ago we came across complex refractive indices inferred from measurements that at first glance seemed unassailable. Scattering of visible (laser) light versus scattering angle by single levitated spheres was fit by a nonlinear least-squares analysis to yield their complex refractive index. What could be more perfect? A single particle, known to be spherical and homogeneous, to which theory is unquestionably applicable. In the paper reporting the imaginary (in more than one sense) refractive indices so obtained, one of the materials was an oil with an imaginary index alleged to be 0.01. This oil came out of a bottle at least 10 cm in diameter. From Eq. (3.123) it follows that the corresponding absorption length of this oil at visible wavelengths is about $5 \mu\text{m}$. Thus we would have to conclude that the oil was poured out of a bottle about 20,000 absorption lengths in diameter. To show that 0.01 is wrong by at least a factor of 10,000 all you have to do is pick up a bottle of the oil and note that you can see through it. What went wrong? In the first place, the experimenter did not know, did not understand, or was incapable of drawing any conclusions from Eq. (3.123), the fundamental equation relating n'' to a more directly measurable and understandable quantity. He also violated the first commandment of measurements: Thou shalt know the magnitude of what you intend to measure before you measure it. Without knowing this magnitude you cannot design a proper method for reliably measuring it. The method under scrutiny here lacked *sensitivity*. Within experimental uncertainty, the angular scattering pattern was the same for imaginary indices of 0.000001 and 0.1 and everything between. So the inferred, but hopelessly incor-

rect, imaginary index 0.01 was the experimental error for a correct value at least 10,000 times smaller.

Misconceptions about imaginary indices are the rule in the atmospheric science literature. They are treated as adjustable parameters, often without any understanding of their magnitudes or how they vary with frequency for different kinds of materials. It is common to find assertions that an imaginary index of 0.01 corresponds to “weak” absorption (at visible wavelengths), presumably because $0.01 \ll 1$. But as we have seen, an imaginary index of 0.01 corresponds to an absorption length of a few micrometers.

An aerosol is a suspension of particles in a gas and includes the gas. A great amount of pointless effort has been expended on determining the refractive index of the atmospheric aerosol even though it does not exist, just as the refractive index of a cow or the Amazon jungle does not exist. What we have here is the failure to distinguish between the properties of a material and of a body. A refractive index is a property of a (homogeneous) material, not of a heterogeneous body. There is no such entity as the complex refractive index of the atmospheric aerosol, an ever changing witch’s brew containing particles of different and generally unknown composition, size, and shape. Individual particles may not even be composed of a single material but are aggregates of particles, like raisins in a loaf of bread coated with sesame seeds.

The complex refractive index of a homogeneous (on the scale of the wavelength) material is a more or less well-defined quantity and can in principle be measured, although in practice this may be difficult. Let’s back up a step and outline how refractive indices are measured by experimenters who know what they are doing. Refractive indices cannot be measured directly but are inferred, usually from measurements of reflection and transmission by homogeneous samples (slabs) of the material of interest. These samples are optically smooth, of uniform and known thickness. The theory of reflection and transmission by such slabs as a function of refractive index is then used to invert measurements to obtain the desired refractive index. Sounds simple. But there are several catches. One is that the theory has to be applicable to the sample. Thus, if the theory is that for specular reflection by a uniformly thick, homogeneous sample, the sample bloody well better conform to this. Moreover, the measured quantities have to change when the refractive index changes. And if the refractive index only reproduces the measurements used to obtain it, it would be of little value. That is, if all we could do is infer a refractive index from measurements of reflectivity and transmissivity, then use it to calculate these same quantities (within experimental error), we’d just be chasing our tails. But properly measured optical constants have a wider validity. For example, they can be used to calculate reflection and transmission by slabs of *any* thickness, or scattering and absorption by homogeneous particles of *any* size and shape (for which a theory exists). This is not tail chasing.

Now consider particles in the atmosphere. In general, their size, shape, and composition are *not* known. They may not be homogeneous and their concentration (number density) varies from point to point, from time to time. Unlike samples prepared in the laboratory, experimenters have no control over the characteristics of atmospheric particles. Nevertheless, one can make various measurements of scattering by them. Problems arise, however, when one attempts to invert these measurements to obtain the (nonexistent) refractive index of the particles, a single complex number at a given wavelength that in some sense is a valid optical property. Between a measurement and a desired quantity to be inferred from it always lies a

theory. What theory is used to invert measurements of scattering by atmospheric particles? Mie theory (see following section), of course, which is valid only for homogeneous spheres, possibly combined with radiative transfer theory for a plane-parallel, uniform medium. An inversion scheme based on this theory will dutifully give results, although as in the example of the oil droplets, they may be hopelessly wrong. All that one can say for sure is that the inferred quantity is consistent, within experimental error, with the theory: if you invert the measurements with Mie theory to obtain refractive indices, then use these refractive indices in Mie theory, you will recover (within error bars) the measurements. All you are doing here is chasing your tail.

Now contrast what folks have been trying to do for particles in the atmosphere with the oil droplet measurements discussed previously. The experimenter had absolute control over his sample: a single, homogeneous sphere, of known size and composition. And yet the error in his imaginary index inferred from measurements was at least a factor of 10,000. Think of how much more difficult it is to measure (remotely) the refractive index of particles, the size distribution, shape, composition, and concentration of which are unknown and over which the experimenter has no control.

Taken in the round, meteorologists don't know much about electromagnetic theory, optics, and solid state physics. But they do know a lot about fluid mechanics (or at least one branch of it). If we were to propose measuring the "effective" viscosity of the ocean-atmosphere system, meteorologists would double over in laughter. Would even the most naive modeler try to model the dynamics of the coupled ocean and atmosphere by treating this system as a single fluid with some kind of bogus effective viscosity? We think not. But these same folks swallow without hesitation the equally bogus notion of an effective refractive index of the atmospheric aerosol.

3.5.2 Scattering by an Isotropic, Homogeneous Sphere

An isotropic, homogeneous sphere is the simplest finite scatterer, the theory of scattering by which is attached to the name of Gustav Mie. So firm is this attachment that in defiance of logic and history every particle under the sun has been dubbed a "Mie scatterer", and Mie scattering has been promoted from a particular theory of limited applicability to the unearned rank of general scattering process.

Mie was not the first to solve the problem of scattering by an arbitrary sphere. It would be more correct to say that he was the last. He gave his solution in recognizably modern notation and also addressed a real problem: the colors of colloidal gold. For these reasons his name is attached to the sphere scattering problem even though he had illustrious predecessors, most notably Lorenz (not to be confused with Lorentz). This is an example in which eponymous recognition has gone to the last discoverer rather than to the first.

Mie scattering is not a physical process; *Mie theory* is one among many. Strictly speaking, it isn't even exact because it is based on continuum electromagnetic theory, itself approximate, and on illumination by a plane wave infinite in lateral extent.

Scattering by a sphere can be determined using various approximations and methods bearing little resemblance to Mie theory: Fraunhofer theory, geometrical optics, anomalous diffraction theory, coupled-dipole, T-matrix method, etc. Thus is a sphere a Mie scatterer or an anomalous diffraction scatterer or a coupled-dipole scatterer? The possibilities are endless.

When a physical process can be described by several different theories, it is inadvisable to attach the name of one of them to it.

There is no distinct boundary between so-called Mie and Rayleigh scatterers. Mie theory includes Rayleigh theory (for spheres), which is a limiting theory strictly applicable only as the size of the particle shrinks to zero. Even for spheres uncritically labeled “Rayleigh spheres”, there are always deviations between the Rayleigh and Mie theories. By hobbling one’s thinking with a supposedly sharp boundary between Rayleigh and Mie scattering, one risks throwing some interesting physics out the window. Whether a particle is a Mie or Rayleigh scatterer is not absolute. A particle may be graduated from Rayleigh to Mie status merely by a change of wavelength of the illumination. One often encounters statements about Mie scattering by cylinders, spheroids, coated spheres and other nonspherical or inhomogeneous particles. Judged historically, these statements are nonsense. Mie never considered any particles other than homogeneous spheres.

Logic would seem to demand that if a particle is a Mie scatterer, then Mie theory can be applied to scattering by it. This fallacious notion has caused and will continue to cause mischief, and is probably the best reason to cease referring to “Mie particles” or “Mie scatterers”. Using Mie theory for scattering by particles other than spheres, especially near the backward direction, is risky.

More often than not, a better term than Mie or Rayleigh scattering is available. If the scatterers are molecules, molecular scattering is better than Rayleigh scattering (itself an imprecise term): the former term refers to an agent, the latter to a theory. Mie scatterer is just a needlessly aristocratic name for a humble sphere. Whenever Mie scatterer is replaced with sphere, the result is clearer. If qualifiers are needed, one can add small or large compared with the wavelength or comparable with the wavelength.

Briefly, the solution to the problem of scattering by an arbitrary homogeneous sphere illuminated by a plane wave can be obtained by expanding the incident, scattered, and internal electric and magnetic fields in series of vector spherical harmonics (general solutions to the equations of the electromagnetic field in spherical coordinates). The coefficients of these expansion functions are chosen so that the tangential components of the fields are continuous across the surface of the sphere. Thus this scattering problem is formally identical to reflection and refraction because of interfaces, although the sphere problem is considerably more complicated because the scattered and internal fields are not plane waves.

Observable quantities are expressed in terms of the complex scattering coefficients a_n and b_n in the expansions of the scattered electric and magnetic fields. For example, the cross sections are infinite series:

$$C_{\text{ext}} = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \Re\{a_n + b_n\}, \quad (3.132)$$

$$C_{\text{sca}} = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \{|a_n|^2 + |b_n|^2\}. \quad (3.133)$$

The scattering coefficients can be written

$$a_n = \frac{[D_n(mx)/m + n/x]\psi_n(x) - \psi_{n-1}(x)}{[D_n(mx)/m + n/x]\xi_n(x) - \xi_{n-1}(x)}, \quad (3.134)$$

$$b_n = \frac{[mD_n(mx) + n/x]\psi_n(x) - \psi_{n-1}(x)}{[mD_n(mx) + n/x]\xi_n(x) - \xi_{n-1}(x)}, \quad (3.135)$$

where ψ_n and ξ_n are Riccati–Bessel functions and the logarithmic derivative is

$$D_n(\rho) = \frac{d}{d\rho} \ln \psi_n(\rho). \quad (3.136)$$

The *size parameter* x is ka , where a is the radius of the sphere and k is the wavenumber of the incident radiation in the surrounding medium (assumed nonabsorbing); m is the complex refractive index (discussed in the previous subsection) of the sphere relative to the (real) refractive index of the surrounding medium. Equations (3.134)–(3.136) are one among many ways of writing the scattering coefficients, some of which are more suited to computations than others. Scattering in any direction for any state of polarization of the incident illumination is also determined by the scattering coefficients.

A good rule of thumb is that the number of terms required for convergence of the series in Eqs. (3.132) and (3.133) is approximately the size parameter ($2\pi a/\lambda$). Raindrops are of order 1 mm, and hence their size parameter at visible wavelengths is of order 10,000. The details of rainbows do emerge from Mie theory but at the cost of summing 10,000 terms, which is why we often resort to approximations such as geometrical optics, which sheds light on some but not all features of rainbows (see Sec. 8.4.2). To describe all their features we have to resort to Mie theory, and even it isn't good enough because raindrops are not spheres (cloud droplets are), and their departures from sphericity have observable consequences. And this leads us to the dubious notion of an “equivalent sphere”, the search for which rivals the quest of alchemists for recipes for transforming base metals into gold. Alas, just as there is no such recipe, there is no such thing as an equivalent sphere, one with all the same scattering and absorbing properties as a non-spherical particle. In the first place, such a particle is defined by what it is *not*, the only characteristic shared by all non-spherical particles. A non-cat is any animal that is not a cat, which leaves us with a menagerie housing everything from elephants and giraffes to porcupines and shrews. Because a collection of randomly oriented non-spherical particles has the same symmetry as a sphere, it is sometimes argued that the (ensemble averaged) scattering properties of the collection are the same as those of suitably chosen “equivalent spheres”. Not true. The error here resides in confusing the symmetry of an ensemble of particles with that of a single particle. No matter what incantation is used for conjuring the properties of an equivalent sphere, differences between scattering and absorption by it and by a non-spherical particle always exist. Sometimes these differences are huge, sometimes not. Beware of equivalent spheres.

During the Great Depression mathematicians were put to work computing tables of trigonometric and other functions. The results of their labors now gather dust in libraries. Today, these tables could be generated more accurately in minutes on a pocket calculator. A similar fate has befallen Mie calculations. Before fast computers were inexpensive, tables of scattering functions for limited ranges of size parameter and refractive index were published. Today, these tables could be generated in minutes on a personal computer. Algorithms are more valuable and enduring than tables of computations, which are mostly useless except as checks for someone developing and testing algorithms. The primary tasks in Mie calculations are computing the functions in Eqs. (3.134) and (3.135) and summing series such as

Eqs. (3.132) and (3.133). Nowadays Mie codes abound. You can find them in books, on the Internet, and probably at upscale supermarket checkout stands.

Cross sections versus radius or wavelength convey physical information; efficiencies versus size parameter convey mathematical information. The size parameter is a variable with less physical content than its components, the whole being less than the sum of its parts. Moreover, cross sections versus size parameter (or its inverse) are not equivalent to cross sections versus wavelength. Except in dreamland, refractive indices vary with wavelength, and the Mie coefficients depend on x and m , wavelength being explicit in the first, implicit in the second.

Particles Much Smaller than the Wavelength

For sufficiently small x and $|m|x$, the volumetric extinction and scattering cross sections for spheres are approximately

$$C_{\text{ext}}/v \approx \frac{6\pi}{\lambda} \Im \left\{ \frac{m^2 - 1}{m^2 + 2} \right\}, \quad (3.137)$$

$$C_{\text{sca}}/v \approx \frac{24\pi^3 v}{\lambda^4} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2, \quad (3.138)$$

where v is the particle volume and λ is the wavelength in the (negligibly absorbing) material surrounding the sphere. Similar equations hold, in particular the volume dependence, for small, homogeneous particles of other shapes. Note the similarity of Eq. (3.138) to Eq. (3.114) for reflection by a thin slab. These equations are the source of a nameless paradox, which is disinterred from time to time, a corpse never allowed to rest in peace. If the sphere is nonabsorbing (m real), Eq. (3.137) yields a vanishing extinction cross section whereas Eq. (3.138) yields a non-vanishing scattering cross section, and yet extinction can never be less than scattering. Both equations were obtained from power series in the size parameter x . Equation (3.137) is the first term in the series. To be consistent, both cross sections must be expanded to the same order in x . When this is done, the paradox vanishes. In fact, it never existed.

Yet another pointless paradox arises from the curious definition of the *radar backscattering cross section* as 4π times the differential scattering cross section in the backward direction. For a small sphere this leads to a backscattering cross section 50% *greater* than its total scattering cross section, which at first glance certainly is cause for head scratching. The *radar reflectivity* is the sum of the radar backscattering cross sections of all the scatterers in a unit volume.

Figure 3.9 shows the scattering and absorption cross sections of a water droplet 20 μm in diameter over six wavelength decades. For wavelengths much greater than the diameter, scattering is a linear function (on a logarithmic plot) of wavelength with slope approximately -4 , in accordance with Eq. (3.138). At these wavelengths, extinction is dominated by absorption, and so according to Eq. (3.137), absorption should decrease (again, on a logarithmic plot) linearly with slope -1 . This is not quite what occurs (see Fig. 3.9) because the complex refractive index also varies with wavelength (Fig. 3.8) in this region.

Although Eq. (3.138) is the scattering cross section of a small particle, it still must contain enough molecules that it can be assigned a refractive index (a molecule cannot). Nevertheless,

let's throw caution to the wind and extrapolate this equation to molecular sizes. If we use the refractive index of water and a particle diameter of 0.3 nm, we obtain a scattering cross section in the middle of the visible spectrum of about $0.7 \times 10^{-19} \mu\text{m}^2$. That for air (an average over all molecules but predominantly nitrogen) is about $4.6 \times 10^{-19} \mu\text{m}^2$. This isn't bad agreement considering that the scattering cross section of water vapor is less than that of air (nitrogen and oxygen), molecular diameters are not well defined, and that the scattering cross section depends on the sixth power of diameter.

Particles Much Larger than the Wavelength

Another paradox has a name: the *extinction paradox*. For a compact particle, such as a sphere, the extinction cross section has the limiting value

$$\lim_{a \rightarrow \infty} C_{\text{ext}} = 2G, \quad (3.139)$$

where a is a linear dimension of the particle (for a sphere its radius) and G its projected geometrical cross sectional area. The factor 2 in Eq. (3.139) has caused people to sweat: according to geometrical optics it should be 1. Equation (3.139) seems to imply that a large (compared with the wavelength) particle is too big for its britches by a factor of two. Geometrical optics, according to which a beam of light is imagined to be a bundle of rays, is reckoned to be a good approximation for objects much larger than the wavelength. Thus every ray that intersects a particle with geometrical area G should be either absorbed or deviated, whereas rays that lie outside this shadow region should pass unscathed. The catch here is that rays don't exist and no matter how large a finite particle is, it always exhibits some departures, possibly small, from geometrical optics, in this instance very close to the forward direction.

We noted previously (and show in Fig. 3.13) that scattering is more peaked in the forward direction the larger the particle. Theory counts scattered light as removed from a (monodirectional) beam no matter how small the scattering angle. To measure the full extinction cross section of an indefinitely large particle would require a detector with vanishingly small acceptance angle. But any real detector necessarily collects some of the near-forward scattered light, which reduces the extinction cross section from its theoretical maximum. When near-forward scattered light is included in the measurement, the limiting extinction cross section [Eq. (3.139)] drops to G , as expected on the basis of intuition molded by geometrical optics. The distinction here is between the real and the ideal. A real detector measures the extinction cross section

$$C_{\text{ext}} = \int_{\text{acc}} \frac{dC_{\text{sca}}}{d\Omega} d\Omega, \quad (3.140)$$

where C_{ext} is the ideal (theoretical) extinction cross section and integration is over the acceptance solid angle of the detector.

In Fig. 3.9 the two asymptotes, G and $2G$, are shown by dotted lines. At sufficiently short wavelengths, where extinction is dominated by scattering, the scattering cross section does indeed approach the asymptote $2G$. But note also that over a range of intermediate wavelengths the scattering and absorption cross sections are each approximately equal to G .

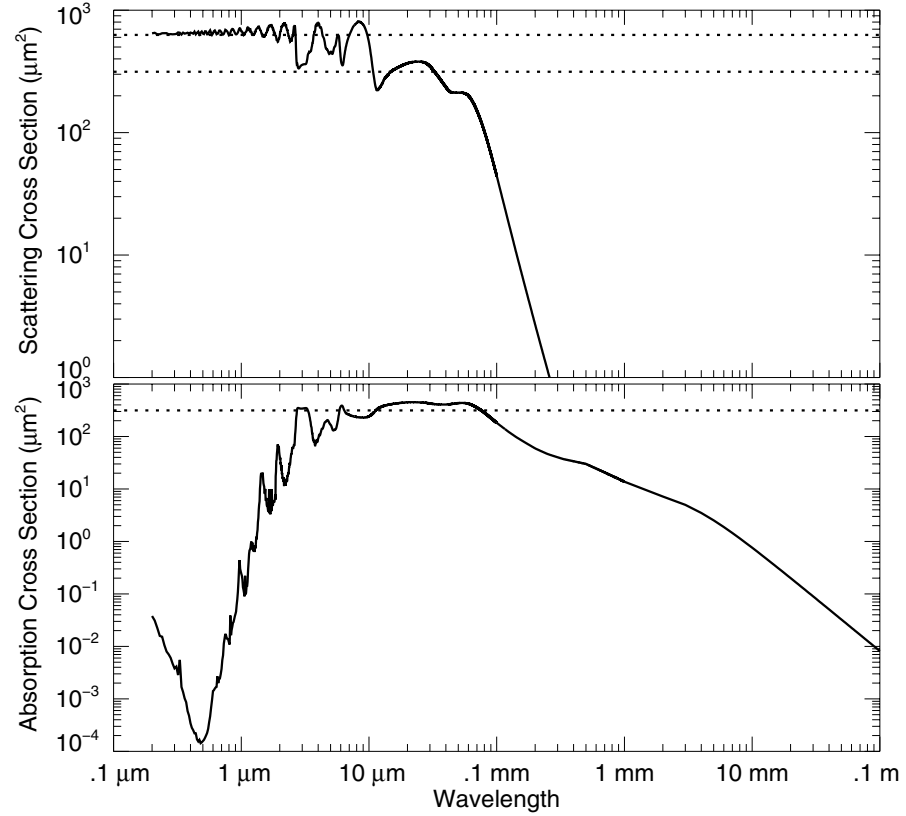


Figure 3.9: Scattering (top panel) and absorption (bottom panel) cross sections for a water droplet of diameter $20\ \mu\text{m}$. The two horizontal dotted lines in the top panel indicate the geometrical cross sectional area and twice this area; the horizontal dotted line in the bottom panel indicates the geometrical cross sectional area.

Those Elusive Small Droplets

We state in Section 1.4.2 that the question of what is *the* average size of a cloud droplet should be greeted with a horselaugh but only sketched the reason why. We can now expand further on this armed with the results of this section.

Cloud droplets are large compared with the wavelengths of visible radiation but small compared with those of microwaves and radar. At visible wavelengths droplet scattering cross sections are proportional to the square of droplet diameter. Extinction is dominated by scattering (Fig. 3.9) and extinction is nearly the asymptotic value [Eq. (3.139)]. At microwave wavelengths, however, extinction is dominated by absorption, and hence from Eq. (3.137), the absorption cross section is proportional to droplet diameter cubed. But at these wavelengths, the scattering cross section, and hence also the radar backscattering cross section, is propor-

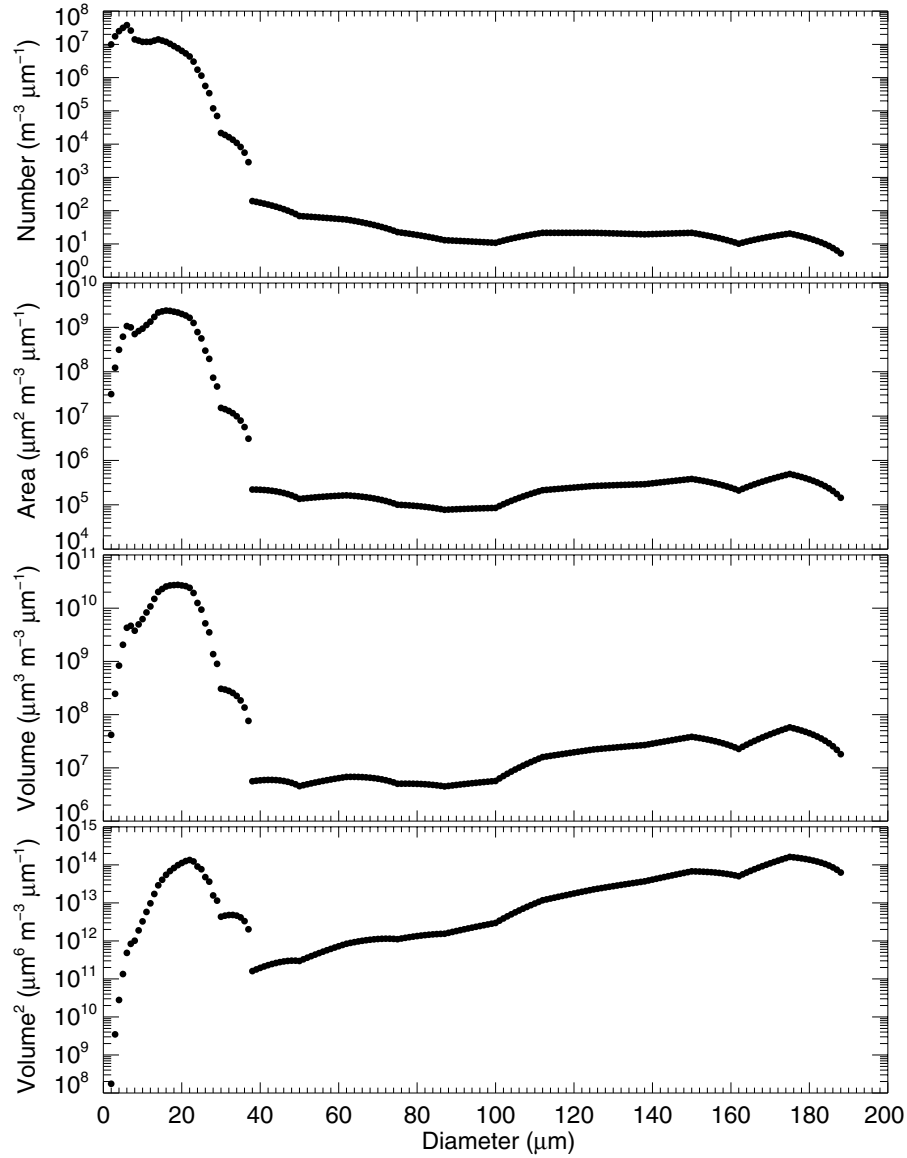


Figure 3.10: The top panel shows an *in situ* measurement of the droplet number size distribution in a stratus cloud. From this distribution the distribution functions for cross sectional area, volume, and volume squared follow.

tional to droplet diameter to the sixth power. These different dependences on size have profound consequences for inferring droplet properties remotely from radiation measurements.

Direct measurements made using an airplane flying through (water) stratus clouds near State College, Pennsylvania are displayed in Fig. 3.10, which shows the number distribution of

droplets, from which area (diameter squared), volume (diameter cubed), and volume squared (diameter to the sixth power) distributions are calculated. The number density distribution (droplets per unit volume) peaks at about $6\text{ }\mu\text{m}$. For diameters greater than about $40\text{ }\mu\text{m}$, the number density is about six decades smaller than at the peak. Scattering of visible radiation peaks at about $16\text{ }\mu\text{m}$, at which diameter the number density is more than a factor of two less than its peak value. The peak in the volume distribution is shifted even further, to about $20\text{ }\mu\text{m}$. And although the peak in the volume squared distribution is shifted to only about $22\text{ }\mu\text{m}$, droplets larger than about $40\text{ }\mu\text{m}$ contribute more to radar backscattering signals than the smaller droplets even though they are a million times more abundant.

Consider how depressing the message Fig. 3.10 conveys to anyone with hopes of inferring droplet size distributions remotely from radiation measurements. The number distribution is greatest for the smaller sizes, whereas properties on which any scheme for remotely inferring sizes might be based are distributed quite differently. For example, if the smallest droplets ($< 6\text{ }\mu\text{m}$) were to be removed entirely from the cloud, the volume, and even more so the volume-squared distributions would hardly miss them. What is needed in order to detect (remotely) the smaller droplets is some electromagnetic property that depends only on number density or, at worst, on diameter. But such a property does not seem to exist.

3.5.3 Some Observable Consequences of Scattering of Visible Radiation by Spherical Particles

The previous section is a panoramic view of scattering and absorption over many wavelength decades. In this section we focus on less than one-third of one of these decades with the aim of helping you to understand what you can see with your own eyes.

Size Dependence

Figure 3.11 shows the volumetric scattering cross section, in the middle of the visible spectrum, of water droplets in air and varying in diameter from about 10 nm to 1 mm . As predicted by simple coherence arguments (Sec. 3.4.8), volumetric scattering increases linearly (on a logarithmic scale) with increasing diameter until a droplet is approximately the size of the wavelength, reaches a peak, oscillates a bit, and decreases linearly (note the similarity to the volumetric reflection cross section in Fig. 3.7). This tremendous variation in volumetric scattering, about four decades, solely as a consequence of the state of aggregation of water molecules, has readily observable consequences. We can see through tens of kilometers of clear, moist air, but when some of the water vapor in it condenses into clouds only tens of meters thick they can obscure the sun. And yet we usually can see through rain shafts, which are born from clouds.

Wavelength Dependence

Now let's turn to the wavelength dependence over the visible spectrum of scattering by spheres of fixed size. Figure 3.12 shows scattering by oil droplets, assumed negligibly absorbing at visible wavelengths, with diameters much smaller than ($0.05\text{ }\mu\text{m}$), comparable with ($0.8\text{ }\mu\text{m}$), and much greater than ($10\text{ }\mu\text{m}$) the wavelengths of visible light. We took the (real) refractive index to be a constant 1.5, which does not introduce serious errors over this limited spectral

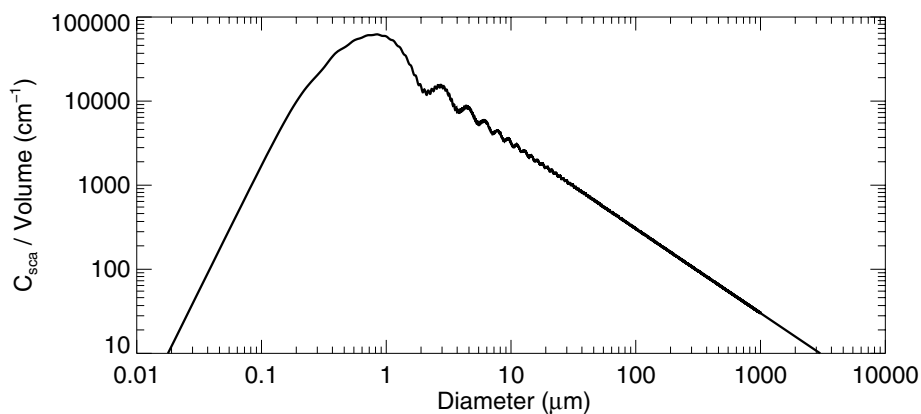


Figure 3.11: Volumetric scattering cross section, at a wavelength of $0.55 \mu\text{m}$, for water droplets with diameters varying over six decades. Except for a proportionality factor, volumetric scattering is scattering per molecule.

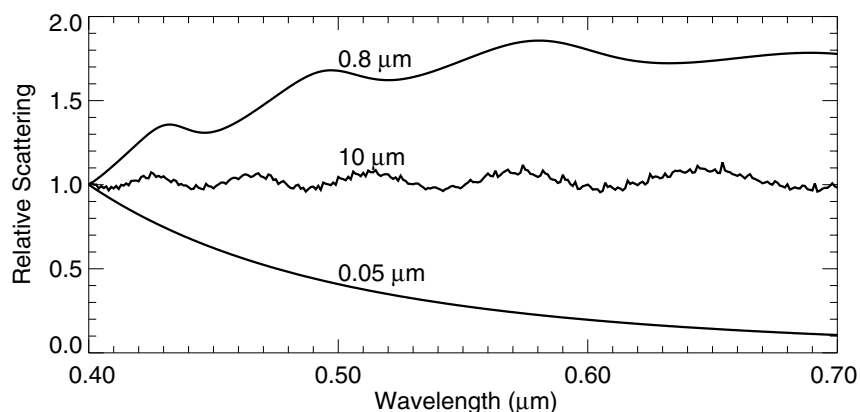


Figure 3.12: Spectral scattering relative to that at $0.4 \mu\text{m}$ for oil droplets ($n = 1.5$) of the three diameters shown on the curves.

range. As expected on the basis of the simple theory of dipolar scattering (Sec. 3.2), scattering by the smallest droplets follows Rayleigh's scattering law, whereas because of non-negligible phase differences, scattering by the largest droplets is nearly independent of wavelength. Note that the spectral dependence of scattering is the same for all droplets smaller than a certain size and larger than a certain size. But between these two extremes lies a comparatively narrow range of droplets that scatter light at the long wavelength end of the spectrum more than at the short end, the inverse of Rayleigh's law.

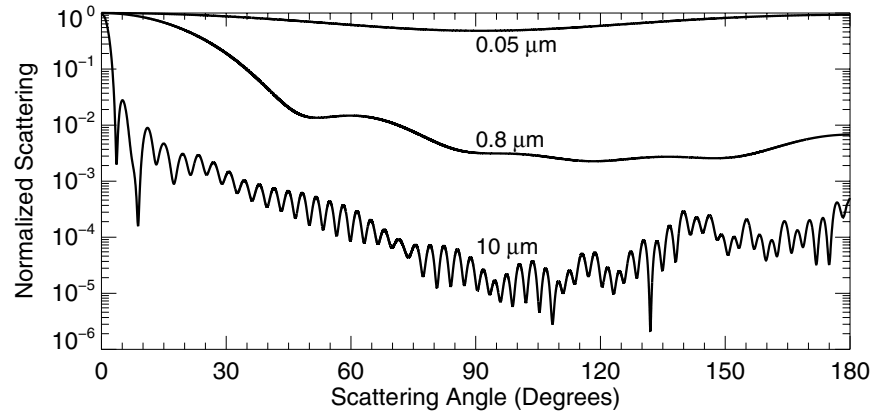


Figure 3.13: Angular dependence of scattering (normalized to 1 in the forward direction) of incident unpolarized light by water droplets of the three diameters shown on the curves. The wavelength of the incident illumination is $0.55 \mu\text{m}$.

Sunlight or moonlight would be selectively (according to wavelength) transmitted by a thin suspension of the smallest droplets. The transmitted light would be reddened, which means that its spectrum would be shifted toward longer wavelengths relative to the incident spectrum, not necessarily that the transmitted light would be perceptually red, whereas the spectrum of light transmitted by the largest droplets would be essentially that of the incident light. Both observations are what we have come to consider the norm. We are not startled when we see a yellow, orange, or red sun or moon. But the spectrum of the light transmitted through a thin veil of oil droplets of intermediate size would be shifted toward shorter wavelengths, yielding a blue or green sun or moon, which would get our attention. This requires a cloud of droplets of just the right size and mostly this size, which is rare in the atmosphere, so rare that it occurs only “once in a blue moon.” Astronomers call the second full moon in a month a blue moon, which has absolutely nothing to do with its color. The last reliable reports of blue and green suns and moons occurred in 1950, and were attributed to oily smoke (which is why we did calculations for oil droplets) from large forest fires in Canada.

Directional Dependence

Because scattering by two or more dipoles is coherent in phase in the forward direction, and only in this direction, we expect scattering by any particle to increase faster with increasing size in this direction than in any other direction. This expectation, based on simple physical reasoning, is supported by calculations (Fig. 3.13) of angular scattering of unpolarized visible light by water droplets with diameters much smaller than ($0.05 \mu\text{m}$), comparable with ($0.8 \mu\text{m}$), and much larger than ($10 \mu\text{m}$) visible wavelengths. The larger the droplet, the more that scattering by it is peaked in the forward direction. Also, the larger the droplet, the more

maxima and minima in its scattering diagram, as expected on the basis of the simple interference arguments in Section 3.4.8.

Observable consequences of strong forward–backward asymmetry in scattering by droplets are not hard to come by, a spider’s web festooned with drops like pearls on a string, for example. A web filament is a long, cylindrical scatterer, for which scattering of sunlight is highly concentrated in one direction. Unless you happen to be looking in that direction, you won’t see webs, which is why you run into them on walks through woods. But if a web has collected fog droplets, you can readily see it, especially if you are looking toward the sun. And then the web almost disappears if you view it with the sun at your back. While driving across the Great Plains of the United States we have watched the greatly changing brightness of spray from center-pivot irrigation as our viewing angle changed.

The wiggles and bumps in Figs. 3.11, 3.12, and especially 3.13 are what might be called *obvious* manifestations of interference. Absorption dampens and may completely iron out such wiggles and bumps, which should hardly come as a surprise: interference filters are not made of strongly absorbing materials. But there is just as much interference in a metallic particle with a smooth scattering diagram as in a negligibly absorbing particle. The scattering pattern for both kinds of particles results from interference between all the waves scattered by their constituent molecules. We have become accustomed to associating interference with wiggles and bumps, but their absence does not necessarily signal no interference.

References and Suggestions for Further Reading

For a biography of Maxwell see Ivan Tolstoy, 1981: *James Clerk Maxwell: A Biography*. University of Chicago. This is a model of scientific biography comprehensible to a non-scientific audience. It is short, well written, and without a single equation conveys the contributions of a scientist whose fame rests mostly on equations. For a much more detailed and technical history of electromagnetism (including electromagnetic radiation) see Edmund Whittaker, 1987: *A History of the Theories of Aether and Electricity. I. The Classical Theories, II. The Modern Theories*. Tomash/American Institute of Physics.

The microscopic (scattering) theory of reflection and refraction is discussed in Max Born and Emil Wolf, 1965: *Principles of Optics*. 3rd rev. ed. Pergamon, pp. 98–108.

A derivation of the radiative reaction is given in John David Jackson, 1975: *Classical Electrodynamics*, 2nd ed., John Wiley & Sons, Ch. 17.

Our treatment of the vibrating string was heavily influenced by Dudley H. Towne, 1967: *Wave Phenomena*, Addison-Wesley, one of the best intermediate-level physics textbooks ever published.

For an expository article on coherence see A. J. Forrester, 1956: On coherence properties of light waves. *American Journal of Physics*, Vol. 24, pp. 192–6.

We criticize, on physical grounds, the usual textbook treatment of the two-slit interference experiment. Tony Rothman, 2003: *Everything’s Relative and Other Fables from Science and*

Technology, John Wiley & Sons, Ch. 2 criticizes it on historical grounds. He casts serious doubts on whether Thomas Young ever did the experiment so often attributed to him.

The simple theory of the diffraction grating is discussed in, for example, Francis A. Jenkins and Harvey E. White, 1976: *Fundamentals of Optics*, 4th ed. McGraw-Hill, Ch. 17. Several papers on grating anomalies, including those by Rayleigh and Wood, are reprinted in Daniel Maystre, Ed., 1992: *Selected Papers on Diffraction Gratings*. SPIE Optical Engineering Press, Sec. 5.

For more on laser speckle see J. C. Dainty, Ed., 1984: *Laser Speckle and Related Phenomena*, 2nd ed., Springer, and M. Francon, 1979: *Laser Speckle and Applications in Optics*, Academic.

For more on optical heterodyning, with atmospheric applications, see James C. Owens, 1969: Optical Doppler measurement of microscale winds. *Proceedings of the IEEE*, Vol. 57, pp. 530–7; Robert T. Menzies, 1976: Laser heterodyne detection techniques, in *Laser Monitoring of the Atmosphere*, E. D. Hinkley, Ed., Springer, pp. 297–353; Rob Frehlich, 1996: Coherent Doppler lidar measurements of winds, in *Trends in Optics: Research, Development and Applications*, Anna Consortini, Ed., Academic, pp. 351–70.

Our treatment of the Doppler effect (Sec. 3.4.6) was inspired by Thomas P. Gill, 1965: *The Doppler Effect: An Introduction to the Theory of the Effect*, Logos Press. p. vii: “One might think of a Doppler effect arising when letters are posted from successive railway stations on a long journey.” Gill is one of the few authors to note that the Doppler effect has nothing fundamental to do with waves, although he adds that “it must be at once admitted that all practical applications of any importance are concerned with periodic processes.”

For a simple demonstration of the Doppler effect, a discussion of its history, and some of its applications, see Craig F. Bohren, 1991: *What Light Through Yonder Window Breaks?*, John Wiley & Sons, Ch. 14.

For a good derivation of the relativistic Doppler effect, including the longitudinal Doppler effect, see Thomas M. Helliwell, 1966: *Introduction to Special Relativity*, Allyn and Bacon, pp. 116–22. This is an excellent little book on relativity. Especially recommended is Appendix B (pp. 148–50), which lists the advantages and disadvantages of a velocity-dependent mass. Contrary to what still seems to be widely believed, a velocity-dependent mass is not an essential part of relativity. To use it or not is a matter of taste, not necessity.

Because of television it is fairly well known that Doppler radar is used to observe weather systems (see, e.g., Louis J. Battan, 1979: *Radar Observation of the Atmosphere*, University of Chicago; Richard J. Doviak and Dusan S. Zrnic, 1993: *Doppler Radar and Weather Observations*, 2nd ed., Academic). Perhaps less well known is an application closer to home: observation of the human (and animal) heart. Both the senior author and one of his dogs, Zas, have had echo cardiograms in which Doppler-shifted ultrasound (high frequency sound waves) is used to measure blood flow in the heart. And both have slight mitral valve leaks,

detected by means of the Doppler effect. As the veterinary cardiologist said upon examining Zas's echo cardiogram, "The apple doesn't fall far from the tree." For an elementary treatment of the use of Doppler ultrasound in medicine see Frederick W. Kremkau, 1998: *Diagnostic Ultrasound: Principles and Instruments*, 5th ed., W. B. Saunders.

An English translation of Einstein's 1910 paper, Theory of the opalescence of homogeneous liquids and mixtures of liquids in the vicinity of the critical state, is in Jerome Alexander, Ed., 1926: *Colloid Chemistry*, Vol. I, Chemical Catalog Co., pp. 323–39. This same collection contains an excellent overview (pp. 340–52) of light scattering by gases, liquids, and solids, W. H. Martin's, The scattering of light in one-phase systems. Although Martin's article is now a bit dated, it is still worth reading because of its clarity and the insights it provides. Martin claims to have been the first to state (in 1913) "based on direct experimental evidence that any dust-free medium scatters light."

What Einstein did without molecules was done with them by Bruno H. Zimm, 1945: Molecular theory of the scattering of light in fluids. *Journal of Chemical Physics*, Vol. 13, pp. 141–5.

For many more details about absorption and scattering by particles see Hendrik C. van de Hulst, 1981: *Light Scattering by Small Particles*. Dover; Diran Deirmendjian, 1969: *Electromagnetic Scattering on Polydispersions*. Elsevier; Milton Kerker, 1969: *The Scattering of Light and Other Electromagnetic Radiation*. Academic; Craig F. Bohren and Donald R. Huffman, 1983: *Absorption and Scattering of Light by Small Particles*. Wiley-Interscience; H. Moysés Nussenzweig, 1992: *Diffraction Effects in Semiclassical Scattering*. Cambridge University Press; Walter T. Grandy, Jr., 2000: *Scattering of Waves from Large Spheres*. Cambridge University Press; Michael I. Mishchenko, Larry D. Travis, and Andrew A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge University Press.

For a brief overview of scattering by particles see Craig F. Bohren, 1995: Scattering by particles in *Handbook of Optics*, Vol. 1, 2nd ed., Michael Bass, Eric W. Van Stryland, David R. Williams, William L. Wolfe, Eds. McGraw-Hill, pp. 6.1–6.21.

For a biographical sketch of Mie see Pedro Lilienfeld, 1991: Gustav Mie: the person. *Applied Optics*, Vol. 30, pp. 4696–8. Pedro was granted a unique opportunity to have a bit of word fun. At the time of publication his affiliation was MIE, Inc. (Measuring Instruments for the Environment), and is so denoted in his paper.

For brevity we call the theory of scattering of a plane wave by a homogeneous sphere Mie theory, but you also will find it called Mie–Debye theory, Lorenz–Mie theory, or even Lorenz–Mie–Debye theory. Once again Stiglitz's law of eponymy (Sec. 2.1) holds. Mie was almost the *last* to solve this scattering problem (Debye's paper was published a year later) but his solution is in recognizably modern notation and he applied it to a real problem, the colors of metals in the colloidal state. For an overview of the contributions of Ludvig Lorenz see Helge Kragh, 1991: Ludvig Lorenz and nineteenth century optical theory: the work of a great Danish scientist. *Applied Optics*, Vol. 30, pp. 4688–95. And for a comprehensive historical

survey see Nelson A. Logan, 1965: Survey of some early studies of the scattering of plane waves by a sphere. *Proceedings IEEE*, Vol. 53, pp. 773–85.

A history of light scattering, From Leonardo to the graser: Light scattering in perspective, was published serially by John D. Hey in *South African Journal of Science*: 1983, Vol. 79, pp. 11–27, 310–24; 1985, Vol. 81, pp. 77–91, 601–13; Vol. 82, pp. 356–60.

Many of the seminal papers on light scattering are reprinted in the compilations by Milton Kerker, Ed., 1988: *Selected Papers on Light Scattering* (2 Vols.), SPIE Optical Engineering Press, 1988 and Philip L. Marston, Ed., 1993: *Selected Papers on Geometrical Aspects of Light Scattering*. SPIE Optical Engineering Press.

For some of the most recent work on scattering by nonspherical particles see Michael I. Mishchenko, Joop W. Hovenier, and Larry D. Travis, Eds., 2000: *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*. Academic.

The diagrams in Figs. 3.11, 3.12, and 3.13 are for spheres, which are *compact* particles, and also negligibly absorbing at the wavelengths of interest. A compact particle is one that more or less fills a continuous region of space (without macroscopic gaps). But not all particles are compact and hence one should be wary about generalizations based on the scattering properties of compact particles. Soot aggregates (produced in fires) are examples of particles that are not compact. Such particles have been called fractal particles, but fractal-like would be more appropriate given that a true fractal is self-similar at *all* scales, whereas soot aggregates and other fractal-like particles are not. For a good review see Chris M. Sorensen, 2001: Light scattering by fractal aggregates: a review. *Aerosol Science and Technology*, Vol. 35, pp. 648–87.

A must read is E. Scott Barr, 1955: Concerning index of refraction and density. *American Journal of Physics*, Vol. 23, 623–4.

It is depressing to realize that as long ago as 1899 it was known that there is a distinction between signal and group velocity, and by 1907 that the signal velocity cannot be greater than c . By 1914 two papers had appeared, back to back, the first by Arnold Sommerfeld, the second by Léon Brillouin, in what was arguably the leading physics journal of the day, in which wave propagation in a dispersive medium was worked out in elaborate detail. English translations of both papers are given as Chapters 2 and 3 in Léon Brillouin's *Wave Propagation and Group Velocity*, 1960, Academic. These chapters require knowledge of the theory of functions of a complex variable, but the introductory chapter does not make great mathematical demands of readers and the introductory section of the second chapter is mostly qualitative. Figure 19 in Chapter 3 deserves careful study.

For a superb compendium of optical constants see Edward D. Palik, Ed., 1998: *Handbook of Optical Constants of Solids* (3 Vols.). Academic. Each volume is divided roughly into two equal sections. The first section describes methods for measuring optical constants and their physical interpretation. The second section is a compilation of measured optical constants for

various materials. Perusal of these volumes makes it abundantly clear that determining optical constants even for well-defined, homogeneous samples of known geometry is difficult, and that much of what is in textbooks about refractive indices is wrong or at best misleading.

A refractive index is an average property. For an average to mean something it has to be taken over a sufficiently large number of values, which is why a single (complex) number describes to good approximation the response of a molecular medium to excitation by radiation of a given frequency. Such a medium (e.g., pure water) is composed of a great many individual scatterers (i.e., molecules) in a volume of one cubic wavelength. A corollary of this is that a medium composed of many particles in a cubic wavelength also can be assigned a refractive index. To the exciting radiation, the particles are large molecules. Determining the effective refractive index of mixtures of molecules and its extension to composite particulate media is nearly 200 years old, the first such problem being that of determining the refractive index of air, a mixture of gases. To good approximation, the refractive index of a mixture of ideal gases is the volume-weighted average of the refractive indices of its components. But this simple mixing rule does not hold as well for liquids or, worse yet, mixtures of particles and over the years many attempts have been made to formulate mixing rules with a greater range of validity. For more on this see Akhlesh Lakhtakia, Ed., 1996: *Selected Papers on Linear Optical Composite Materials*, SPIE Optical Engineering Press. This compendium contains many of the classical papers on the subject of homogenizing composite media.

Spectral measurements of a blue sun observed at Edinburgh in 1950 are reported by Robert Wilson, 1951: The blue sun of 1950 September. *Monthly Notices of the Royal Astronomical Society*, Vol. 111, pp. 478–89, reprinted in the compendium *Selected Papers on Scattering in the Atmosphere* cited at the end of Chapter 8.

Problems

3.1. If multiple scattering is negligible, a beam of radiation is attenuated exponentially by scattering just as it is by absorption (see Sec. 5.2.3). Thus the irradiance of a monodirectional beam transmitted through a scattering medium (with negligible absorption) is $F = F_0 \exp(-\beta z)$, where β is the scattering coefficient. As in Section 2.3, we can determine the negative of the flux divergence: $\beta F_0 \exp(-\beta z)$. But this leads to a perplexing result: the negative flux divergence and hence the heating rate appear to be nonzero even in a medium with no absorption; moreover, this heating rate (at $z = 0$) increases with increasing scattering coefficient. Explain.

3.2. In Section 3.3 we derive a wave equation for the (small) displacements of a string under tension. The motion of the string is not damped. Generalize this one-dimensional wave equation to a string with damping. Use the simplest damping law you can think of. After you have obtained the wave equation, find harmonic solutions (sinusoidal waves). Interpret your solutions. Find the power transmitted by the waves. Interpret your result. This problem is much easier if you use complex solutions to the wave equation.

3.3. Several years ago we received a letter from a former student. He had seen a series of dark bands in the sky, which he interpreted as an “interference pattern”. To the extent that we

understood the student's reasoning, it seemed that he attributed this pattern to interference of the light scattered by ice particles in a cirrus cloud. Suppose that you had received this letter. What would have been your response?

3.4. Why does the wave function $A \exp(ikx - i\omega t)$ not strictly describe a laser beam?

HINT: This is not a difficult problem. You do not need to know anything about lasers but you should have at least seen one. You may assume that light from a laser is exactly monochromatic and exactly monodirectional. This is not a problem about lasers but rather about reality, idealizations of reality, and the differences between them.

3.5. The following appeared in a paper (Sean Twomey, 1980: Direct visual photometric technique for estimating absorption in collected aerosol samples. *Applied Optics*, Vol. 19, pp. 1740–1) in a scientific journal several years ago:

“There is at the present time considerable interest in absorption of solar radiation by aerosol particles in the atmosphere, and it has been speculated that such absorption may have climatic significance.

Ground-based measurements of solar radiation are fundamentally incapable of giving such absorption directly and without assumptions, since . . .”

Please supply the missing arguments in support of the author's assertion in the second paragraph (italicized). Before you write a long dissertation, reflect for a moment that the author of this paper needed only 17 words with which to make his point forcefully and clearly.

3.6. Estimate an upper limit for the imaginary part of the refractive index of ordinary window glass in the middle of the visible spectrum. Then do the same for tinted glass.

3.7. Sir William Henry Bragg wrote two delightful works of popular science, one entitled *The Universe of Light*, the other *The World of Sound*. Discuss the difference in the titles.

3.8. The frequency ν of incident light scattered by a moving scatterer is Doppler shifted to a frequency ν' . But the energy of a photon is $h\nu$, which means that the energy of a photon is changed upon scattering. Doesn't this violate conservation of energy? Suppose that a (nonabsorbing) gas is illuminated by light. What happens to the energy of the gas as a consequence of the Doppler shift?

HINT: It might help to review Section 1.1.

3.9. These days it has become fairly common in the United States to be driving along a highway and come upon a sign at the side of the road flashing YOUR SPEED IS X MPH. Is this speed equal to, greater than, or less than your actual speed? Explain your answer. Do not invoke “experimental error” or “friction” or some other such nonsense.

3.10. In H. G. Wells's novel, *The Invisible Man*, the title character makes himself invisible by drinking a potion that makes his refractive index equal to that of air at visible wavelengths. This then enables him to do all kinds of dastardly things that he would not have been able to do if he were visible. Setting aside the fact that drinking something that would transform the refractive index of all human cells to 1 is absurd, everything else in this novel is scientifically correct with the exception of one subtle but fatal flaw. What is it? The Invisible Man functions just like ordinary human beings.

3.11. We state in Section 1.1 that frequency is more fundamental than wavelength but did not explain why. Supply the missing explanation.

3.12. Suppose that a blackbody with temperature T is moving relative to a stationary observer with line-of-sight velocity v . Derive an expression for the difference between the frequency of blackbody peak irradiance measured by the stationary observer and that measured by an observer moving with the blackbody. Don't bother to use the relativistically correct expression for the Doppler shift. The approximate expression, Eq. (3.74), is accurate (within about 10%) for speeds less than about $0.5c$.

HINT: You can use results in Section 1.2 or simply write down the answer. What does this result tell you about the velocity required such that the perceived color of a (visible) blackbody (e.g., a star) is appreciably different for the two different observers? This problem has historical significance. Doppler's original work was aimed at explaining the different observed colors of stars by their different speeds. You might revisit this problem after reading Section 4.3.

3.13. We state in Section 3.4.6 that the distribution function Eq. (3.78) for molecular velocity components along a particular direction is a member of the same family as the Maxwell-Boltzmann distribution for molecular kinetic energies. Show that this distribution function, Eq. (1.8), follows from Eq. (3.78).

HINT: For a gas in equilibrium the probability distributions for velocity components in three orthogonal directions must be the same and independent. Transform from the probability distribution in rectangular Cartesian coordinates (in velocity space) to the probability distribution in spherical polar coordinates, which then leads to the probability distribution for kinetic energies by the theorem for transforming variables of integration.

3.14. A refractive index is an average quantity that describes, to good approximation (often but not always), the electromagnetic response of systems composed of huge numbers of scatterers (e.g., atoms and molecules). This is possible only because one cubic wavelength of the media to which refractive indices usually are applied (solids and liquids even at ultraviolet wavelengths) contain a great many scatterers, and hence an average over the relevant volume means something. But to a wave of sufficiently long wavelength, a cloud droplet (or a snow grain) is a molecule insofar as its scattering properties are concerned (i.e., a dipole scatterer). For what wavelengths is it therefore not absurd to speak of the refractive index of an ordinary cloud? What about for a snowpack (snow on the ground)?

3.15. We found the following example of the Doppler effect "in every day life" on a web site devoted to explaining the principles of Doppler echocardiography: "an observer stationed on a highway overpass readily notices that the pitch of the sound made from the engine of a passing automobile changes from high to low as the car approaches and then travels into the distance. The engine is emitting the same sound as it passes beneath, but the observer notices a change in pitch dependent on the speed of the auto and its direction." Sounds good but we don't believe it. Why?

3.16. Equation (3.111) for normal-incidence reflection was given for a negligibly absorbing medium (real refractive index). But this equation is valid even if the refractive index is complex. Under what general conditions (magnitudes of the real and imaginary parts of the complex refractive index) is this reflectivity close to 1?

3.17. A particle is a collection of molecules, and hence scattering by a particle is the sum of scattering by all its constituent molecules. Suppose that there were no such thing as in-

terference of light. What would the scattering diagram (scattered irradiance as a function of direction) of a particle look like?

3.18. Is the scattering cross section at visible wavelengths of a small (compared with the wavelength) water droplet equal to, greater than, or less than the scattering cross section of a bubble of the same size in water? The source of illumination is the same for both scatterers. Try to answer this by physical reasoning before attempting a calculation. If there is a difference, try to give a physical explanation.

3.19. Fog lamps on automobiles often are yellow. Perhaps there are sound reasons for this. Be that as it may, various reasons have been given. Here is one from a website that answers questions about science. “It is important for fog lights to be one [color] (rather than white) because the different wavelengths (colors) of visible light scatter off the fog droplets differently. This phenomenon is known as ‘dispersion’, because the different colors of light in an image will separate from each other. . . If you illuminate the road with only one wavelength (color) of light, the images of the objects you will see still become somewhat blurry because of the scattering of the light by the fog, but at least you won’t have the extra problems from dispersion. So, if we want to use just one wavelength of light, which wavelength should we use? It turns out that light with short wavelengths scatters more than light with long wavelengths. So a long wavelength is best.” Discuss this explanation in light of what is in this chapter and what you know about fog (as the term is used by meteorologists and most people). We revisit this problem in Chapter 4 (see Prob. 4.64).

3.20. A plane harmonic wave transmitted in an optically homogeneous medium is the superposition of an incident (exciting) wave in free space and all the forward-scattered waves it excites. Thus the phase of the transmitted wave is, in general, shifted from that of the incident wave. Show that the (real) refractive index is a phase-shift parameter. This is a simple problem to reveal another – and in our opinion better – interpretation of the refractive index than the ratio of the speed of light in a vacuum to the phase speed in a medium. Phase speeds have no physical significance whereas phase shifts do. Moreover, once we look upon the refractive index as a phase shift parameter, this should put an end to hand-wringing over supposedly forbidden ($> c$) “speeds of light” because we are not encumbered by expectations that transmitted waves must always be retarded or advanced.

3.21. In his contribution to *Colloid Chemistry* (cited in the references) W. H. Martin (p. 549) asserts that “From the standpoint of the diffraction theory a perfectly regular arrangement of molecules within a crystal would result in no scattering”. By “scattering” here he means lateral scattering. Examine this assertion by determining scattering by a one-dimensional crystal: N identical dipolar scatterers equally spaced along a line. The direction of the incident wave is parallel to this line. Follow the same approach as that in Section 3.4.9. Equation (1.39) is useful. The expression you obtain for the scattered energy might look familiar. Consider the limiting cases of a separation between adjacent scatterers much smaller than and comparable with the wavelength. Carefully examine the strict validity of every assumption and approximation you make, especially when the separation is much smaller than the wavelength. This is an open-ended problem that you can explore almost endlessly. Its purpose is to make you think about the perils of making absolute statements (“no scattering”) on the basis of approximate theories, especially if the magnitude of the physical quantity to be predicted is small in some sense.

3.22. In Section 3.4.6 we show that the frequency of radiation shifts because of interaction with moving matter. But this means that photon energies also shift, which seems to suggest that energy is not conserved. Show that it is. For simplicity, take the moving matter to be a disc of area A , much larger than the wavelength, moving with a speed v much smaller than c . The disc is illuminated at normal incidence and is nonabsorbing with reflectivity R and transmissivity T .

3.23. Consider two stretched strings, both infinitely long, with different mass densities, and joined at $x = 0$. The tension in both strings is, of course, the same. A time-harmonic wave propagating in the $+x$ direction and incident on this interface between two different strings gives rise to a reflected wave in the string $-\infty < x < 0$ and a transmitted wave in the string $0 < x < \infty$. At $x = 0$ the string displacement y and its slope must be continuous. Determine the amplitudes of the reflected and transmitted waves (relative to that of the incident wave). Show that energy flux (time-averaged) is conserved. This problem is essentially identical to reflection and transmission of electromagnetic waves. Note the similarity between your result and the Fresnel coefficients in Section 7.2.

3.24. This problem is related to Problem 3.23. Consider a finite stretched string of length h , attached at one end ($x = 0$) to an infinite string with a different mass density and at the other end ($x = h$) to another infinite string with the same mass density as the string $-\infty < x < 0$; an incident harmonic wave in this region propagates in the $+x$ direction. Find the ratio of the reflected ($x < 0$) energy flux (time-averaged) to the incident flux (reflectivity) and the ratio of the transmitted ($x > h$) energy flux to the incident flux (transmissivity). Compare your result with Eq. (5.24). Show that energy is conserved (reflectivity + transmissivity = 1).

HINT: Two sets of boundary conditions now have to be satisfied (at $x = 0$ and $x = h$).

3.25. We show in Section 2.8 that the absorption cross section per unit volume of a large (compared with the wavelength), weakly absorbing particle is approximately equal to the bulk absorption coefficient of the particle material. Show that this same approximation holds (to within about a factor of 2) at the other extreme, a particle small compared with the wavelength, provided that the real part of the refractive index lies between about 1 and 2 and the imaginary part is sufficiently small. Take the particle to be a sphere in air. This result is a useful rule of thumb if used with due caution.

3.26. How does the scattering cross section of atmospheric molecules compare with their geometrical cross section? Keep in mind that the diameter of a molecule is not a precisely defined quantity.

3.27. According to Eqs. (3.134) and (3.135) the extinction and scattering cross sections for a small sphere (in air, say) become infinite if the complex refractive index has a particular (finite) value. What is that value? Can you find materials that, at some wavelengths, have approximately this value?

3.28. Derive the Doppler line shape [Eq. (2.135)] beginning with Eqs. (3.79) and (3.80). Keep in mind that the Doppler shift for absorption is half that for scattering.

3.29. The scattering cross section of a particle is often greater in air than in a negligibly absorbing medium with a higher refractive index (e.g., water). But is this *always* true? Can you find any exceptions?

HINTS: First try to answer this question in a general way for an arbitrary sphere, which for simplicity take to be negligibly absorbing. Keep in mind that the wavelength for the size parameter is that in the medium surrounding the particle. The chain rule for differentiation is useful. To obtain a specific example, consider a sphere very small compared with the wavelength.

3.30. Give the simplest example you can think of to show that the scattering cross section of a particle has no necessary relation to its mass.

3.31. Consider two dipolar antennas each driven to radiate by their own independent power sources. Absorption is negligible. The antennas are mutually coherent and their power sources are fixed. Move the antennas closer and closer together until the radiation from them is in phase. Now the total radiated power should be twice what it was when the antennas were far apart (relative to the wavelength). But this seems to violate conservation of energy because the power sources are fixed. Explain.

3.32. Many years ago one of the authors used to give lectures as part of a short course on atmospheric aerosols. One of his topics was optical constants. He would pull out of his pocket a chunk of obsidian, a black glass of volcanic origin, wave it in the air and, with lots of dramatic flourishes, proclaim it to be the blackest substance on the planet, “blackier than the inside of a bruised crow”, and assert that it therefore *must* have a huge imaginary index. Then he’d ask his audience to estimate its value. These were seasoned engineers and scientists, not innocent students. As expected (and hoped for) almost everyone would fall into the trap and blurt out idiotic responses: ten, one hundred, one thousand, a million. One of those who estimated a million earned his living (predictably) from remotely measuring the imaginary index of the atmospheric aerosol. But a very few clever folks didn’t fall into the trap and were able to come up with a reasonable estimate. How?

3.33. We note in Section 1.1 that photons carry linear momentum, and hence a beam of radiation can exert a force (radiation pressure) on objects (see Prob. 1.32). The force exerted by a monodirectional beam on a particle is proportional to the beam’s momentum flux, which is proportional to its irradiance. This force is the product of the momentum flux and the radiation pressure cross section. A rigorous derivation of this cross section is difficult, even for a sphere, but by physical grounds you should be able to write down the correct expression for an arbitrary particle.

We also note in Section 1.1 that photons can carry angular momentum, a consequence of which is that a circularly polarized beam (Sec. 7.1.3) can exert a *radiation torque* on a particle. This torque is the product of the angular momentum flux (proportional to irradiance) and the radiation torque cross section. A rigorous derivation of this cross section for a sphere is even more formidable than a derivation of the radiation pressure cross section. See Philip L. Marston and James H. Crichton, 1984: Radiation torque on a sphere caused by a circularly-polarized electromagnetic wave, *Physical Review A*, Vol. 30, pp. 2508–16. The senior author of this book was the second reviewer of this paper. The first reviewer had rejected it on the grounds that the result was obvious. No, the result was simple and plausible, which is not the same as obvious.