Oscillations and Waves

Richard Fitzpatrick

Professor of Physics
The University of Texas at Austin

Contents

1 Introduction **5**

2 Simple Harmonic Oscillation **7**
2.1 Mass on a Spring 7
2.2 Simple Harmonic Oscillator Equation 12
2.3 LC Circuit 14
2.4 Simple Pendulum 18
2.5 Exercises 20

3 Damped and Driven Harmonic Oscillation **23**
3.1 Damped Harmonic Oscillation 23
3.2 Quality Factor 26
3.3 LCR Circuit 28
3.4 Driven Damped Harmonic Oscillation 29
3.5 Driven LCR Circuit 34
3.6 Transient Oscillator Response 36
3.7 Exercises 40

4 Coupled Oscillations **43**
4.1 Two Spring-Coupled Masses 43
4.2 Two Coupled LC Circuits 48
4.3 Three Spring Coupled Masses 51
4.4 Exercises 53

5 Transverse Standing Waves **55**
- 5.1 Normal Modes of a Beaded String . . . 55
- 5.2 Normal Modes of a Uniform String . . . 63
- 5.3 General Time Evolution of a Uniform String . . . 69
- 5.4 Exercises . . . 76

6 Longitudinal Standing Waves **81**
- 6.1 Spring Coupled Masses . . . 81
- 6.2 Sound Waves in an Elastic Solid . . . 86
- 6.3 Sound Waves in an Ideal Gas . . . 91
- 6.4 Fourier Analysis . . . 94
- 6.5 Exercises . . . 100

7 Traveling Waves **103**
- 7.1 Standing Waves in a Finite Continuous Medium . . . 103
- 7.2 Traveling Waves in an Infinite Continuous Medium . . . 104
- 7.3 Wave Interference . . . 107
- 7.4 Energy Conservation . . . 109
- 7.5 Transmission Lines . . . 113
- 7.6 Reflection and Transmission at Boundaries . . . 115
- 7.7 Electromagnetic Waves . . . 122
- 7.8 Exercises . . . 127

8 Wave Pulses **131**
- 8.1 Fourier Transforms . . . 131
- 8.2 General Solution of the Wave Equation . . . 137
- 8.3 Bandwidth . . . 142
- 8.4 Exercises . . . 148

9 Dispersive Waves **151**
- 9.1 Pulse Propagation . . . 151
- 9.2 Electromagnetic Wave Propagation in Plasmas . . . 154
- 9.3 Electromagnetic Wave Propagation in Conductors . . . 161
- 9.4 Surface Wave Propagation in Water . . . 165
- 9.5 Exercises . . . 170

10 Multi-Dimensional Waves **173**
- 10.1 Plane Waves . . . 173
- 10.2 Three-Dimensional Wave Equation . . . 175
- 10.3 Laws of Geometric Optics . . . 175

10.4 Waveguides 178
10.5 Cylindrical Waves 180
10.6 Exercises 181

11 Wave Optics 183
11.1 Introduction 183
11.2 Two-Slit Interference 183
11.3 Coherence 191
11.4 Multi-Slit Interference 197
11.5 Fourier Optics 202
11.6 Single-Slit Diffraction 203
11.7 Multi-Slit Diffraction 206
11.8 Exercises 208

12 Wave Mechanics 213
12.1 Introduction 213
12.2 Photoelectric Effect 214
12.3 Electron Diffraction 216
12.4 Representation of Waves via Complex Numbers 216
12.5 Schrödinger's Equation 219
12.6 Probability Interpretation of the Wavefunction 220
12.7 Wave Packets 223
12.8 Heisenberg's Uncertainty Principle 227
12.9 Collapse of the Wavefunction 228
12.10 Stationary States 230
12.11 Three-Dimensional Wave Mechanics 234
12.12 Particle in a Finite Potential Well 237
12.13 Square Potential Barrier 241
12.14 Exercises 243

A Useful Information 247
A.1 Physical Constants 247
A.2 Trigonometric Identities 247
A.3 Calculus 248
A.4 Power Series 248

1 Introduction

Oscillations and waves are ubiquitous phenomena that are encountered in many different areas of physics. An *oscillation* is a disturbance in a physical system that is *repetitive in time*. A *wave* is a disturbance in an extended physical system that is both *repetitive in time* and *periodic in space*. In general, an oscillation involves a continuous back and forth flow of energy between two different energy types: *e.g.*, kinetic and potential energy, in the case of a pendulum. A wave involves similar repetitive energy flows to an oscillation, but, in addition, is capable of transmitting energy and information from place to place. Now, although sound waves and electromagnetic waves, for example, rely on quite distinct physical mechanisms, they, nevertheless, share many common properties. The same is true of different types of oscillation. It turns out that the common factor linking various types of wave is that they are all described by the *same* mathematical equations. Again, the same is true of various types of oscillation.

The aim of this course is to develop a unified mathematical theory of oscillations and waves in physical systems. Examples will be drawn from the dynamics of discrete mechanical systems; continuous gases, fluids, and elastic solids; electronic circuits; electromagnetic waves; and quantum mechanical systems.

This course assumes a basic familiarity with the laws of physics, such as might be obtained from a two-semester introductory college-level survey course. Students are also assumed to be familiar with standard mathematics, up to and including trigonometry, linear algebra, differential calculus, integral calculus, ordinary differential equations, partial differential equations, and Fourier series.

The textbooks which were consulted most often during the development of the course material are:

Waves, Berkeley Physics Course, Vol. 3, F.S. Crawford, Jr. (McGraw-Hill, New York NY, 1968).

Vibrations and Waves, A.P. French (W.W. Norton & Co., New York NY, 1971).

Introduction to Wave Phenomena, A. Hirose, and K.E. Lonngren (John Wiley & Sons, New York NY, 1985).

The Physics of Vibrations and Waves, 5th Edition, H.J. Pain (John Wiley & Sons, Chichester UK, 1999).

2 Simple Harmonic Oscillation

2.1 Mass on a Spring

Consider a compact mass m which slides over a frictionless horizontal surface. Suppose that the mass is attached to one end of a light horizontal spring whose other end is anchored in an immovable wall. See Figure 2.1. At time t, let $x(t)$ be the extension of the spring: *i.e.*, the difference between the spring's actual length and its unstretched length. Obviously, $x(t)$ can also be used as a coordinate to determine the instantaneous horizontal displacement of the mass.

The equilibrium state of the system corresponds to the situation in which the mass is at rest, and the spring is unextended (*i.e.*, $x = \dot{x} = 0$, where $\dot{} \equiv d/dt$). In this state, zero horizontal force acts on the mass, and so there is no reason for it to start to move. However, if the system is perturbed from its equilibrium state (*i.e.*, if the mass is displaced, so that the spring becomes extended) then the mass experiences a horizontal *restoring force* given by *Hooke's law*:

$$f(x) = -k\,x. \tag{2.1}$$

Here, $k > 0$ is the so-called *force constant* of the spring. The negative sign indicates that $f(x)$ is indeed a restoring force (*i.e.*, if the displacement is positive then the force is negative, and *vice versa*). Note that the magnitude of the restoring force is *directly proportional* to the displacement of the mass from its equilibrium position (*i.e.*, $|f| \propto x$). Of course, Hooke's law only holds for relatively *small* spring extensions. Hence, the displacement of the mass cannot be made too large. Incidentally, the motion of this particular dynamical system is representative of the motion of a wide variety of mechanical systems when they are *slightly* disturbed from a stable equilibrium state (see Section 2.4).

Newton's second law of motion gives following time evolution equation for the system:

$$m\,\ddot{x} = -k\,x, \tag{2.2}$$

where $\ddot{} \equiv d^2/dt^2$. This *differential equation* is known as the *simple harmonic oscillator equation*, and its solution has been known for centuries. In fact, the solution is

$$x(t) = a\,\cos(\omega\,t - \phi), \tag{2.3}$$

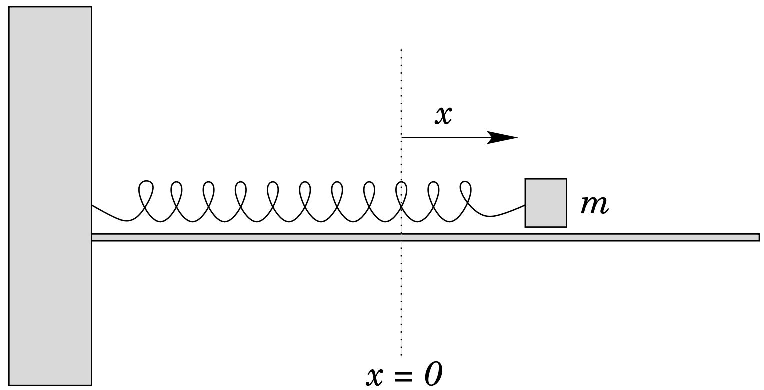


Figure 2.1: *Mass on a spring*

where $a > 0$, $\omega > 0$, and ϕ are constants. We can demonstrate that Equation (2.3) is indeed a solution of Equation (2.2) by direct substitution. Plugging the right-hand side of (2.3) into Equation (2.2), and recalling from standard calculus that $d(\cos\theta)/d\theta = -\sin\theta$ and $d(\sin\theta)/d\theta = \cos\theta$, so that $\dot{x} = -\omega\, a\, \sin(\omega\, t - \phi)$ and $\ddot{x} = -\omega^2\, a\, \cos(\omega\, t - \phi)$, where use has been made of the *chain rule*, we obtain

$$-\, m\, \omega^2\, a\, \cos(\omega\, t - \phi) = -k\, a\, \cos(\omega\, t - \phi). \qquad (2.4)$$

It follows that Equation (2.3) is the correct solution provided

$$\omega = \sqrt{\frac{k}{m}}. \qquad (2.5)$$

Figure 2.2 shows a graph of x versus t obtained from Equation (2.3). The type of behavior shown here is called *simple harmonic oscillation*. It can be seen that the displacement x *oscillates* between $x = -a$ and $x = +a$. Here, a is termed the *amplitude* of the oscillation. Moreover, the motion is *repetitive* in time (*i.e.*, it repeats exactly after a certain time period has elapsed). In fact, the *repetition period* is

$$T = \frac{2\pi}{\omega}. \qquad (2.6)$$

This result is easily obtained from Equation (2.3) by noting that $\cos\theta$ is a periodic function of θ with period 2π: *i.e.*, $\cos(\theta + 2\pi) \equiv \cos\theta$. It follows that the motion repeats every time $\omega\, t$ increases by 2π: *i.e.*, every time t

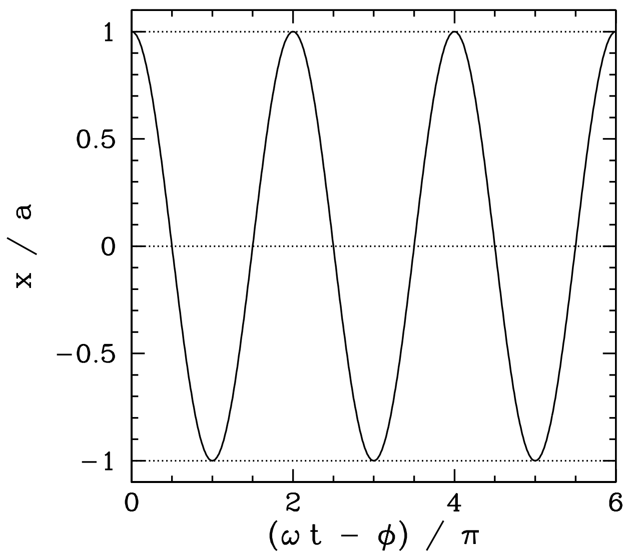


Figure 2.2: *Simple harmonic oscillation.*

increases by $2\pi/\omega$. The *frequency* of the motion (*i.e.*, the number of oscillations completed per second) is

$$f = \frac{1}{T} = \frac{\omega}{2\pi}. \tag{2.7}$$

It can be seen that ω is the motion's *angular frequency*; *i.e.*, the frequency f converted into radians per second. Of course, f is measured in *Hertz*—otherwise known as *cycles per second*. Finally, the *phase angle*, ϕ, determines the times at which the oscillation attains its maximum displacement, $x = a$. In fact, since the maxima of $\cos\theta$ occur at $\theta = n\,2\pi$, where n is an arbitrary integer, the times of maximum displacement are

$$t_{max} = T\left(n + \frac{\phi}{2\pi}\right). \tag{2.8}$$

Clearly, varying the phase angle simply shifts the pattern of oscillation backward and forward in time. See Figure 2.3.

Table 2.1 lists the displacement, velocity, and acceleration of the mass at various different phases of the simple harmonic oscillation cycle. The information contained in this table can easily be derived from Equation (2.3). Note that all of the non-zero values shown in this table represent either the

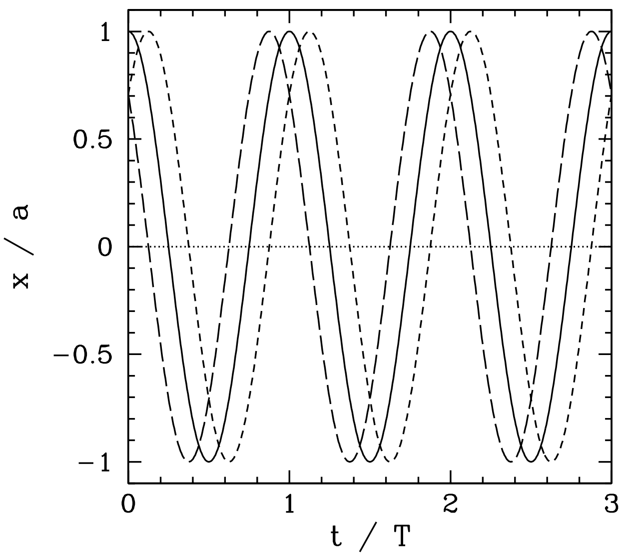


Figure 2.3: *Simple harmonic oscillation. The solid, short-dashed, and long dashed-curves correspond to* $\phi = 0$, $+\pi/4$, *and* $-\pi/4$, *respectively.*

maximum or the minimum value taken by the quantity in question during the oscillation cycle.

We have seen that when a mass on a spring is disturbed it executes *simple harmonic oscillation* about its equilibrium position. In physical terms, if the mass's initial displacement is positive ($x > 0$) then the restoring force is negative, and pulls the mass toward the equilibrium point ($x = 0$). However, when the mass reaches this point it is moving, and its inertia thus carries it onward, so that it acquires a negative displacement ($x < 0$). The restoring force then becomes positive, and again pulls the mass toward the equilibrium point. However, inertia again carries it past this point, and the mass acquires a positive displacement. The motion subsequently repeats it-

$\omega\, t - \phi$	0	$\pi/2$	π	$3\pi/2$
x	$+a$	0	$-a$	0
$\dot{x}$	0	$-\omega\, a$	0	$+\omega\, a$
$\ddot{x}$	$-\omega^2\, a$	0	$+\omega^2\, a$	0

Table 2.1: *Simple harmonic oscillation.*

self *ad infinitum*. The angular frequency of the oscillation is determined by the spring stiffness, k, and the system inertia, m, via Equation (2.5). On the other hand, the amplitude and phase angle of the oscillation are determined by the *initial conditions*. To be more exact, suppose that the instantaneous displacement and velocity of the mass at $t = 0$ are x_0 and v_0, respectively. It follows from Equation (2.3) that

$$x_0 = x(t = 0) = a\,\cos\phi, \tag{2.9}$$

$$v_0 = \dot{x}(t = 0) = a\,\omega\,\sin\phi. \tag{2.10}$$

Here, use has been made of the trigonometric identities $\cos(-\theta) \equiv \cos\theta$ and $\sin(-\theta) \equiv -\sin\theta$. Hence, we deduce that

$$a = \sqrt{x_0^2 + (v_0/\omega)^2}, \tag{2.11}$$

and

$$\phi = \tan^{-1}\left(\frac{v_0}{\omega\,x_0}\right), \tag{2.12}$$

since $\sin^2\theta + \cos^2\theta \equiv 1$ and $\tan\theta \equiv \sin\theta/\cos\theta$.

The kinetic energy of the system, which is the same as the kinetic energy of the mass, is written

$$K = \frac{1}{2}\,m\,\dot{x}^2 = \frac{1}{2}\,m\,a^2\,\omega^2\,\sin^2(\omega\,t - \phi). \tag{2.13}$$

The potential energy of the system, which is the same as the potential energy of the spring, takes the form

$$U = \frac{1}{2}\,k\,x^2 = \frac{1}{2}\,k\,a^2\,\cos^2(\omega\,t - \phi). \tag{2.14}$$

Hence, the total energy is

$$E = K + U = \frac{1}{2}\,k\,a^2 = \frac{1}{2}\,m\,\omega^2\,a^2, \tag{2.15}$$

since $m\,\omega^2 = k$ and $\sin^2\theta + \cos^2\theta \equiv 1$. Note that the total energy is a *constant of the motion*. Moreover, the energy is proportional to the *amplitude squared* of the oscillation. It is clear, from the above expressions, that the simple harmonic oscillation of a mass on a spring is characterized by a continuous backward and forward flow of energy between kinetic and potential components. The kinetic energy attains its maximum value, and the potential energy its minimum value, when the displacement is zero (*i.e.*, when

$x = 0$). Likewise, the potential energy attains its maximum value, and the kinetic energy its minimum value, when the displacement is maximal (*i.e.*, when $x = \pm a$). Note that the minimum value of K is *zero*, since the system is instantaneously at rest when the displacement is maximal.

2.2 Simple Harmonic Oscillator Equation

Suppose that a physical system possesing *one degree of freedom—i.e.*, a system whose instantaneous state at time t is fully described by a *single dependent variable*, $s(t)$—obeys the following time evolution equation [*cf.*, Equation (2.2)]:

$$\ddot{s} + \omega^2\,s = 0, \tag{2.16}$$

where $\omega > 0$ is a constant. As we have seen, this differential equation is called the *simple harmonic oscillator equation*, and has the following solution

$$s(t) = a\,\cos(\omega\,t - \phi), \tag{2.17}$$

where $a > 0$ and ϕ are constants. Moreover, the above equation describes a type of oscillation characterized by a constant *amplitude*, a, and a constant *angular frequency*, ω. The *phase angle*, ϕ, determines the times at which the oscillation attains its maximum value. Finally, the frequency of the oscillation (in Hertz) is $f = \omega/2\pi$, and the period is $T = 2\pi/\omega$. Note that the frequency and period of the oscillation are both determined by the constant ω, which appears in the simple harmonic oscillator equation, whereas the amplitude, a, and phase angle, ϕ, are both determined by the *initial conditions*—see Equations (2.9)–(2.12). In fact, a and ϕ are the two *constants of integration* of the *second-order ordinary differential equation* (2.16). Recall, from standard differential equation theory, that the most general solution of an nth-order ordinary differential equation (*i.e.*, an equation involving a single independent variable and a single dependent variable in which the highest derivative of the dependent with respect to the independent variable is nth-order, and the lowest zeroth-order) involves n arbitrary constants of integration. (Essentially, this is because we have to integrate the equation n times with respect to the independent variable in order to reduce it to zeroth-order, and so obtain the solution, and each integration introduces an arbitrary constant: *e.g.*, the integral of $\dot{s} = a$, where a is a known constant, is $s = a\,t + b$, where b is an arbitrary constant.)

Multiplying Equation (2.16) by $\dot{s}$, we obtain

$$\dot{s}\,\ddot{s} + \omega^2\,\dot{s}\,s = 0. \tag{2.18}$$

However, this can also be written in the form

$$\frac{d}{dt}\left(\frac{1}{2}\,\dot{s}^{\,2}\right)+\frac{d}{dt}\left(\frac{1}{2}\,\omega^2\,s^2\right)=0, \tag{2.19}$$

or

$$\frac{d\mathcal{E}}{dt}=0, \tag{2.20}$$

where

$$\mathcal{E}=\frac{1}{2}\,\dot{s}^{\,2}+\frac{1}{2}\,\omega^2\,s^2. \tag{2.21}$$

Clearly, $\mathcal{E}$ is a *conserved quantity*: *i.e.*, it does not vary with time. In fact, this quantity is generally proportional to the overall energy of the system. For instance, $\mathcal{E}$ would be the energy divided by the mass in the mass-spring system discussed in Section 2.1. Note that $\mathcal{E}$ is either zero or positive, since neither of the terms on the right-hand side of Equation (2.21) can be negative. Let us search for an *equilibrium state*. Such a state is characterized by s = constant, so that $\dot{s}=\ddot{s}=0$. It follows from (2.16) that $s=0$, and from (2.21) that $\mathcal{E}=0$. We conclude that the system can only remain permanently at rest when $\mathcal{E}=0$. Conversely, the system can never permanently come to rest when $\mathcal{E}>0$, and must, therefore, keep moving for ever. Furthermore, since the equilibrium state is characterized by $s=0$, it follows that s represents a kind of "displacement" of the system from this state. It is also apparent, from (2.21), that s attains it maximum value when $\dot{s}=0$. In fact,

$$s_{\rm max}=\frac{\sqrt{2\,\mathcal{E}}}{\omega}. \tag{2.22}$$

This, of course, is the amplitude of the oscillation: *i.e.*, $s_{\rm max}=a$. Likewise, $\dot{s}$ attains its maximum value when $x=0$, and

$$\dot{s}_{\rm max}=\sqrt{2\,\mathcal{E}}. \tag{2.23}$$

Note that the simple harmonic oscillation (2.17) can also be written in the form

$$s(t)=A\,\cos(\omega\,t)+B\,\sin(\omega\,t), \tag{2.24}$$

where $A=a\,\cos\phi$ and $B=a\,\sin\phi$. Here, we have employed the trigonometric identity $\cos(x-y)\equiv\cos x\,\cos y+\sin x\,\sin y$. Alternatively, (2.17) can be written

$$s(t)=a\,\sin(\omega\,t-\phi'), \tag{2.25}$$

where $\phi'=\phi-\pi/2$, and use has been made of the trigonometric identity $\cos\theta\equiv\sin(\pi/2+\theta)$. Clearly, there are many different ways of representing a simple harmonic oscillation, but they all involve *linear* combinations of sine and cosine functions whose arguments take the form $\omega\,t+c$, where c is some constant. Note, however, that, whatever form it takes, a *general* solution to the simple harmonic oscillator equation must always contain *two* arbitrary constants: *i.e.*, A and B in (2.24) or a and ϕ' in (2.25).

The simple harmonic oscillator equation, (2.16), is a *linear* differential equation, which means that if $s(t)$ is a solution then so is $a\,s(t)$, where a is an arbitrary constant. This can be verified by multiplying the equation by a, and then making use of the fact that $a\,d^2s/dt^2 = d^2(a\,s)/dt^2$. Now, linear differential equations have a very important and useful property: *i.e.*, their solutions are *superposable*. This means that if $s_1(t)$ is a solution to Equation (2.16), so that

$$\ddot{s}_1 = -\omega^2\,s_1, \tag{2.26}$$

and $s_2(t)$ is a different solution, so that

$$\ddot{s}_2 = -\omega^2\,s_2, \tag{2.27}$$

then $s_1(t) + s_2(t)$ is also a solution. This can be verified by adding the previous two equations, and making use of the fact that $d^2s_1/dt^2 + d^2s_2/dt^2 = d^2(s_1+s_2)/dt^2$. Furthermore, it is easily demonstrated that *any linear combination* of s_1 and s_2, such as $a\,s_1 + b\,s_2$, where a and b are constants, is a solution. It is very helpful to know this fact. For instance, the solution to the simple harmonic oscillator equation (2.16) with the initial conditions $s(0) = 1$ and $\dot{s}(0) = 0$ is easily shown to be

$$s_1(t) = \cos(\omega\,t). \tag{2.28}$$

Likewise, the solution with the initial conditions $s(0) = 0$ and $\dot{s}(0) = 1$ is clearly

$$s_2(t) = \omega^{-1}\,\sin(\omega\,t). \tag{2.29}$$

Thus, since the solutions to the simple harmonic oscillator equation are superposable, the solution with the initial conditions $s(0) = s_0$ and $\dot{s}(0) = \dot{s}_0$ is $s(t) = s_0\,s_1(t) + \dot{s}_0\,s_2(t)$, or

$$s(t) = s_0\,\cos(\omega\,t) + \frac{\dot{s}_0}{\omega}\,\sin(\omega\,t). \tag{2.30}$$

2.3 LC Circuit

Consider an electrical circuit consisting of an inductor, of inductance L, connected in series with a capacitor, of capacitance C. See Figure 2.4. Such a circuit is known as an LC circuit, for obvious reasons. Suppose that $I(t)$ is the instantaneous current flowing around the circuit. According to standard electrical circuit theory, the potential difference across the inductor is $L\,\dot{I}$. Again, from standard electrical circuit theory, the potential difference across the capacitor is $V = Q/C$, where Q is the charge stored on the capacitor's positive plate. However, since electric charge is *conserved*, the current flowing around the circuit is equal to the rate at which charge accumulates on the capacitor's positive plate: *i.e.*, $I = \dot{Q}$. Now, according to *Kichhoff's second circuital law*, the sum of the potential differences across the various components of a closed circuit loop is equal to zero. In other words,

$$L\,\dot{I} + Q/C = 0. \tag{2.31}$$

Dividing by L, and differentiating with respect to t, we obtain

$$\ddot{I} + \omega^2\,I = 0, \tag{2.32}$$

where

$$\omega = \frac{1}{\sqrt{L\,C}}. \tag{2.33}$$

Comparison with Equation (2.16) reveals that (2.32) is a *simple harmonic oscillator equation* with the associated angular oscillation frequency ω. We conclude that the current in an LC circuit executes simple harmonic oscillations of the form

$$I(t) = I_0\,\cos(\omega\,t - \phi), \tag{2.34}$$

where $I_0 > 0$ and ϕ are constants. Now, according to Equation (2.31), the potential difference, $V = Q/C$, across the capacitor is minus that across the inductor, so that $V = -L\,\dot{I}$, giving

$$V(t) = \sqrt{\frac{L}{C}}\,I_0\,\sin(\omega\,t - \phi) = \sqrt{\frac{L}{C}}\,I_0\,\cos(\omega\,t - \phi - \pi/2). \tag{2.35}$$

Here, use has been made of the trigonometric identity $\sin\theta \equiv \cos(\theta - \pi/2)$. It follows that the voltage in an LC circuit oscillates at the *same frequency* as the current, but with a *phase shift* of $\pi/2$. In other words, the voltage is maximal when the current is zero, and *vice versa*. The amplitude of the voltage oscillation is that of the current oscillation multiplied by $\sqrt{L/C}$. Thus, we can also write

$$V(t) = \sqrt{\frac{L}{C}}\,I(t - \omega^{-1}\,\pi/2). \tag{2.36}$$

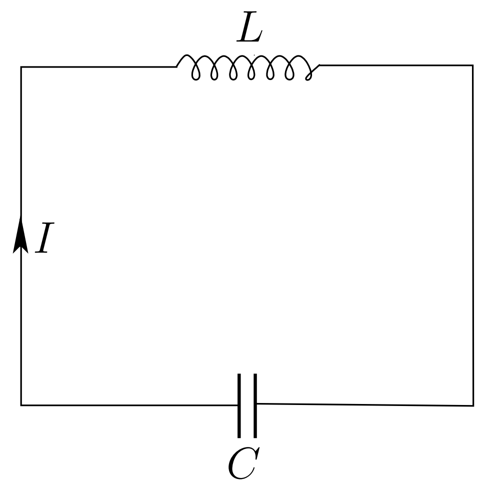


Figure 2.4: *An* LC *circuit.*

Comparing with Equation (2.21), it is clear that

$$\mathcal{E} = \frac{1}{2}\,\dot{I}^{\,2} + \frac{1}{2}\,\omega^2\,I^2 \tag{2.37}$$

is a conserved quantity. However, $\omega^2 = 1/L\,C$, and $\dot{I} = -V/L$. Thus, multiplying the above expression by $C\,L^2$, we obtain

$$E = \frac{1}{2}\,C\,V^2 + \frac{1}{2}\,L\,I^2. \tag{2.38}$$

The first and second terms on the right-hand side of the above expression can be recognized as the instantaneous energies stored in the capacitor and the inductor, respectively. The former energy is stored in the electric field generated when the capacitor is charged, whereas the latter is stored in the magnetic field induced when current flows through the inductor. It follows that (2.38) is the total energy of the circuit, and that this energy is a *conserved quantity*. Clearly, the oscillations of an LC circuit can be understood as a cyclic interchange between electric energy stored in the capacitor and magnetic energy stored in the inductor, much as the oscillations of the mass-spring system studied in Section 2.1 can be understood as a cyclic interchange between kinetic energy stored by the mass and potential energy stored by the spring.

Suppose that at $t = 0$ the capacitor is charged to a voltage V_0, and there is no current flowing through the inductor. In other words, the initial state is one in which all of the circuit energy resides in the capacitor. The initial conditions are $V(0) = -L\,\dot{I}(0) = V_0$ and $I(0) = 0$. It is easily demonstrated that the current evolves in time as

$$I(t) = -\frac{V_0}{\sqrt{L/C}}\,\sin(\omega\,t). \tag{2.39}$$

Suppose that at $t = 0$ the capacitor is fully discharged, and there is a current I_0 flowing through the inductor. In other words, the initial state is one in which all of the circuit energy resides in the inductor. The initial conditions are $V(0) = -L\,\dot{I}(0) = 0$ and $I(0) = I_0$. It is easily demonstrated that the current evolves in time as

$$I(t) = I_0\,\cos(\omega\,t). \tag{2.40}$$

Suppose, finally, that at $t = 0$ the capacitor is charged to a voltage V_0, and the current flowing through the inductor is I_0. Since the solutions of the simple harmonic oscillator equation are superposable, it is clear that the current evolves in time as

$$I(t) = -\frac{V_0}{\sqrt{L/C}}\,\sin(\omega\,t) + I_0\,\cos(\omega\,t). \tag{2.41}$$

Furthermore, it follows from Equation (2.36) that the voltage evolves in time as

$$V(t) = -V_0\,\sin(\omega\,t - \pi/2) + \sqrt{\frac{L}{C}}\,I_0\,\cos(\omega\,t - \pi/2), \tag{2.42}$$

or

$$V(t) = V_0\,\cos(\omega\,t) + \sqrt{\frac{L}{C}}\,I_0\,\sin(\omega\,t). \tag{2.43}$$

Here, use has been made of the trigonometric identities $\sin(\theta - \pi/2) \equiv -\cos\theta$ and $\cos(\theta - \pi/2) \equiv \sin\theta$.

The instantaneous electrical power absorption by the capacitor, which can easily be shown to be minus the instantaneous power absorption by the inductor, is

$$P(t) = I(t)\,V(t) = I_0\,V_0\,\cos(2\,\omega\,t) + \frac{1}{2}\left(I_0^2\sqrt{\frac{L}{C}} - \frac{V_0^2}{\sqrt{L/C}}\right)\sin(2\,\omega\,t), \tag{2.44}$$

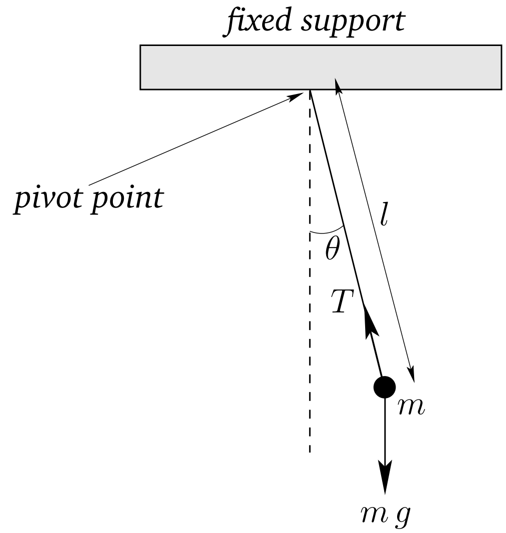


Figure 2.5: *A simple pendulum.*

where use has been made of Equations (2.41) and (2.43), as well as the trigonometric identities $\cos(2\,\theta) \equiv \cos^2\theta - \sin^2\theta$ and $\sin(2\,\theta) \equiv 2\,\sin\theta\,\cos\theta$. Hence, the average power absorption during a cycle of the oscillation,

$$\langle P\rangle = \frac{1}{T}\int_0^T P(t)\,dt, \tag{2.45}$$

is *zero*, since it is easily demonstrated that $\langle\cos(2\,\omega\,t)\rangle = \langle\sin(2\,\omega\,t)\rangle = 0$. In other words, any energy which the capacitor absorbs from the circuit during one part of the oscillation cycle is returned to the circuit without loss during another. The same goes for the inductor.

2.4 Simple Pendulum

Consider a compact mass m suspended from a light inextensible string of length l, such that the mass is free to swing from side to side in a vertical plane, as shown in Figure 2.5. This setup is known as a *simple pendulum*. Let θ be the angle subtended between the string and the downward vertical. Obviously, the stable equilibrium state of the system corresponds to the situation in which the mass is stationary, and hangs vertically down (*i.e.*, $\theta = \dot{\theta} = 0$). The angular equation of motion of the pendulum is simply

$$I\,\ddot{\theta} = \tau, \tag{2.46}$$

where I is the moment of inertia of the mass, and τ the torque acting about the suspension point. For the case in hand, given that the mass is essentially a point particle, and is situated a distance l from the axis of rotation (*i.e.*, the suspension point), it is easily seen that $I = m\,l^2$.

The two forces acting on the mass are the downward gravitational force, $m\,g$, where g is the acceleration due to gravity, and the tension, T, in the string. Note, however, that the tension makes no contribution to the torque, since its line of action clearly passes through the suspension point. From elementary trigonometry, the line of action of the gravitational force passes a perpendicular distance $l\,\sin\theta$ from the suspension point. Hence, the magnitude of the gravitational torque is $m\,g\,l\,\sin\theta$. Moreover, the gravitational torque is a *restoring torque*: *i.e.*, if the mass is displaced slightly from its equilibrium position (*i.e.*, $\theta = 0$) then the gravitational torque clearly acts to push the mass back towards that position. Thus, we can write

$$\tau = -m\,g\,l\,\sin\theta. \tag{2.47}$$

Combining the previous two equations, we obtain the following angular equation of motion of the pendulum:

$$l\,\ddot{\theta} + g\,\sin\theta = 0. \tag{2.48}$$

Note that, unlike all of the other time evolution equations which we have examined so far in this chapter, the above equation is *nonlinear* [since $\sin(a\,\theta) \neq a\,\sin\theta$], which means that it is generally very difficult to solve.

Suppose, however, that the system does not stray very far from its equilibrium position ($\theta = 0$). If this is the case then we can expand $\sin\theta$ in a Taylor series about $\theta = 0$. We obtain

$$\sin\theta = \theta - \frac{\theta^3}{6} + \frac{\theta^5}{120} + \mathcal{O}(\theta^7). \tag{2.49}$$

Clearly, if $|\theta|$ is sufficiently small then the series is dominated by its first term, and we can write $\sin\theta \simeq \theta$. This is known as the *small angle approximation*. Making use of this approximation, the equation of motion (2.48) simplifies to

$$\ddot{\theta} + \omega^2\,\theta \simeq 0, \tag{2.50}$$

where

$$\omega = \sqrt{\frac{g}{l}}. \tag{2.51}$$

Of course, (2.50) is just the *simple harmonic oscillator equation*. Hence, we can immediately write its solution in the form

$$\theta(t) = \theta_0 \cos(\omega\,t - \phi), \tag{2.52}$$

where $\theta_0 > 0$ and ϕ are constants. We conclude that the pendulum swings back and forth at a fixed frequency, ω, which depends on l and g, but is *independent* of the amplitude, θ_0, of the motion. Of course, this result only holds as long as the small angle approximation remains valid. It turns out that $\sin\theta \simeq \theta$ is a good approximation provided $|\theta| \lesssim 6^\circ$. Hence, the period of a simple pendulum is only amplitude independent when the amplitude of the motion is less than about 6°.

2.5 Exercises

1. A mass stands on a platform which executes simple harmonic oscillation in a vertical direction at a frequency of 5 Hz. Show that the mass loses contact with the platform when the displacement exceeds 10^{-2} m.

2. A small body rests on a horizontal diaphragm of a loudspeaker which is supplies with an alternating current of constant amplitude but variable frequency. If diaphragm executes simple harmonic oscillation in the vertical direction of amplitude 10 μm, at all frequencies, find the greatest frequency for which the small body stays in contact with the diaphragm.

3. Two light springs have spring constants k_1 and k_2, respectively, and are used in a vertical orientation to support an object of mass m. Show that the angular frequency of small amplitude oscillations about the equilibrium state is $[(k_1 + k_2)/m]^{1/2}$ if the springs are in parallel, and $[k_1\,k_2/(k_1 + k_2)\,m]^{1/2}$ if the springs are in series.

4. A body of uniform cross-sectional area A and mass density ρ floats in a liquid of density ρ_0 (where $\rho < \rho_0$), and at equilibrium displaces a volume V. Making use of Archimedes principle (that the buoyancy force acting on a partially submerged body is equal to the mass of the displaced liquid), show that the period of small amplitude oscillations about the equilibrium position is
$$T = 2\pi\sqrt{\frac{V}{g\,A}}.$$

5. A particle of mass m slides in a frictionless semi-circular depression in the ground of radius R. Find the angular frequency of small amplitude oscillations about the particle's equilibrium position, assuming that the oscillations are essentially one dimensional, so that the particle passes through the lowest point of the depression during each oscillation cycle.

6. If a thin wire is twisted through an angle θ then a restoring torque $\tau = -k\,\theta$ develops, where $k > 0$ is known as the *torsional force constant*. Consider a so-called *torsional pendulum*, which consists of a horizontal disk of mass M, and moment of inertia I, suspended at its center from a thin vertical wire of negligible mass and length l, whose other end is attached to a fixed support. The disk is free to rotate about a vertical axis passing through the suspension point, but such rotation twists the wire. Find the frequency of torsional oscillations of the disk about its equilibrium position.

7. Suppose that a hole is drilled through a laminar (*i.e.*, flat) object of mass M, which is then suspended in a frictionless manner from a horizontal axis passing through the hole, such that it is free to rotate in a vertical plane. Suppose that the moment of inertia of the object about the axis is I, and that the distance of the hole from the object's center of mass is d. Find the frequency of small angle oscillations of the object about its equilibrium position. Hence, find the frequency of small angle oscillations of a *compound pendulum* consisting of a uniform rod of mass M and length l suspended vertically from a horizontal axis passing through one of its ends.

8. A pendulum consists of a uniform circular disk of radius r which is free to turn about a horizontal axis perpendicular to its plane. Find the position of the axis for which the periodic time is a minimum.

9. A particle of mass m executes one-dimensional simple harmonic oscillation under the action of a conservative force such that its instantaneous x coordinate is
$$x(t) = a\,\cos(\omega\,t - \phi).$$
Find the average values of x, x^2, $\dot{x}$, and $\dot{x}^2$ over a single cycle of the oscillation. Find the average values of the kinetic and potential energies of the particle over a single cycle of the oscillation.

10. A particle executes two-dimensional simple harmonic oscillation such that its instantaneous coordinates in the x-y plane are
$$\begin{aligned} x(t) &= a\,\cos(\omega\,t), \\ y(t) &= a\,\cos(\omega\,t - \phi). \end{aligned}$$
Describe the motion when (a) $\phi = 0$, (b) $\phi = \pi/2$, and (c) $\phi = -\pi/2$. In each case, plot the trajectory of the particle in the x-y plane.

11. An LC circuit is such that at $t = 0$ the capacitor is uncharged and a current I_0 flows through the inductor. Find an expression for the charge Q stored on the positive plate of the capacitor as a function of time.

12. A simple pendulum of mass m and length l is such that $\theta(0) = 0$ and $\dot{\theta}(0) = \omega_0$. Find the subsequent motion, $\theta(t)$, assuming that its amplitude remains small. Suppose, instead, that $\theta(0) = \theta_0$ and $\dot{\theta}(0) = 0$. Find the subsequent motion. Suppose, finally, that $\theta(0) = \theta_0$ and $\dot{\theta}(0) = \omega_0$. Find the subsequent motion.

13. Demonstrate that

$$E = \frac{1}{2}\,m\,l^2\,\dot{\theta}^{\,2} + m\,g\,l\,(\cos\theta - 1)$$

is a constant of the motion of a simple pendulum whose time evolution equation is given by (2.48). (Do not make the small angle approximation.) Hence, show that the amplitude of the motion, θ_0, can be written

$$\theta_0 = 2\,\sin^{-1}\left(\frac{E}{2\,m\,g\,l}\right)^{1/2}.$$

Finally, demonstrate that the period of the motion is determined by

$$\frac{T}{T_0} = \frac{1}{\pi}\int_0^{\theta_0}\frac{d\theta}{\sqrt{\sin^2(\theta_0/2) - \sin^2(\theta/2)}},$$

where T_0 is the period of small angle oscillations. Verify that $T/T_0 \to 1$ as $\theta_0 \to 0$. Does the period increase, or decrease, as the amplitude of the motion increases?

3 Damped and Driven Harmonic Oscillation

3.1 Damped Harmonic Oscillation

In the previous chapter, we encountered a number of energy conserving physical systems which exhibit *simple harmonic oscillation* about a stable equilibrium state. One of the main features of such oscillation is that, once excited, it never dies away. However, the majority of the oscillatory systems which we generally encounter in everyday life suffer some sort of irreversible energy loss due, for instance, to frictional or viscous heat generation whilst they are oscillating. We would therefore expect oscillations excited in such systems to eventually be damped away. Let us examine an example of a damped oscillatory system.

Consider the mass-spring system investigated in Section 2.1. Suppose that, as it slides over the horizontal surface, the mass is subject to a *frictional damping force* which opposes its motion, and is directly proportional to its instantaneous velocity. It follows that the net force acting on the mass when its instantaneous displacement is $x(t)$ takes the form

$$f = -k\,x - m\,\nu\,\dot{x}, \tag{3.1}$$

where $m > 0$ is the mass, $k > 0$ the spring force constant, and $\nu > 0$ a constant (with the dimensions of angular frequency) which parameterizes the strength of the damping. The time evolution equation of the system thus becomes [*cf.*, Equation (2.2)]

$$\ddot{x} + \nu\,\dot{x} + \omega_0^2\,x = 0, \tag{3.2}$$

where $\omega_0 = \sqrt{k/m}$ is the undamped oscillation frequency [*cf.*, Equation (2.5)]. We shall refer to the above as the *damped harmonic oscillator equation*.

Let us search for a solution to Equation (3.2) of the form

$$x(t) = a\,e^{-\gamma\,t}\cos(\omega_1\,t - \phi), \tag{3.3}$$

where $a > 0$, $\gamma > 0$, $\omega_1 > 0$, and ϕ are all constants. By analogy with the discussion in Section 2.1, we can interpret the above solution as a periodic oscillation, of fixed angular frequency ω_1 and phase angle ϕ, whose amplitude *decays exponentially* in time as $a(t) = a\,\exp(-\gamma\,t)$. So, (3.3) certainly seems like a plausible solution for a damped oscillatory system. It is easily demonstrated that

$$\dot{x} = -\gamma\, a\, e^{-\gamma\, t} \cos(\omega_1\, t - \phi) - \omega_1\, a\, e^{-\gamma\, t} \sin(\omega_1\, t - \phi), \tag{3.4}$$

$$\ddot{x} = (\gamma^2 - \omega_1^2)\, a\, e^{-\gamma\, t} \cos(\omega_1\, t - \phi) + 2\,\gamma\,\omega_1\, a\, e^{-\gamma\, t} \sin(\omega_1\, t - \phi), \tag{3.5}$$

so Equation (3.2) becomes

$$\begin{aligned} 0 &= \left[(\gamma^2 - \omega_1^2) - \nu\,\gamma + \omega_0^2\right] a\, e^{-\gamma\, t} \cos(\omega_1\, t - \phi) \\ &\quad + [2\,\gamma\,\omega_1 - \nu\,\omega_1]\, a\, e^{-\gamma\, t} \sin(\omega_1\, t - \phi). \end{aligned} \tag{3.6}$$

Now, the only way in which the above equation can be satisfied at all times is if the coefficients of $\exp(-\gamma\, t)\, \cos(\omega_1\, t - \phi)$ and $\exp(-\gamma\, t)\, \sin(\omega_1\, t - \phi)$ separately equate to zero, so that

$$(\gamma^2 - \omega_1^2) - \nu\,\gamma + \omega_0^2 = 0, \tag{3.7}$$

$$2\,\gamma\,\omega_1 - \nu\,\omega_1 = 0. \tag{3.8}$$

These equations can easily be solved to give

$$\gamma = \nu/2, \tag{3.9}$$

$$\omega_1 = (\omega_0^2 - \nu^2/4)^{1/2}. \tag{3.10}$$

Thus, the solution to the damped harmonic oscillator equation is written

$$x(t) = a\, e^{-\nu\, t/2} \cos\left(\omega_1\, t - \phi\right), \tag{3.11}$$

assuming that $\nu < 2\,\omega_0$ (since $\omega_1^2 = \omega_0^2 - \nu^2/4$ clearly cannot be negative). We conclude that the effect of a relatively small amount of damping, parameterized by the *damping constant* ν, on a system which exhibits simple harmonic oscillation about a stable equilibrium state is to *reduce the angular frequency* of the oscillation from its undamped value ω_0 to $(\omega_0^2 - \nu^2/4)^{1/2}$, and to cause the amplitude of the oscillation to *decay exponentially in time* at the rate $\nu/2$. This modified type of oscillation, which we shall refer to as *damped harmonic oscillation*, is illustrated in Figure 3.1. [Here, $T_0 = 2\pi/\omega_0$, $\nu\, T_0 = 0.5$, and $\phi = 0$. The solid line shows $x(t)/a$, whereas the dashed lines show $\pm a(t)/a$.] Incidentally, if the damping is sufficiently large that $\nu \geq 2\,\omega_0$, which we shall assume is *not* the case, then the system does not

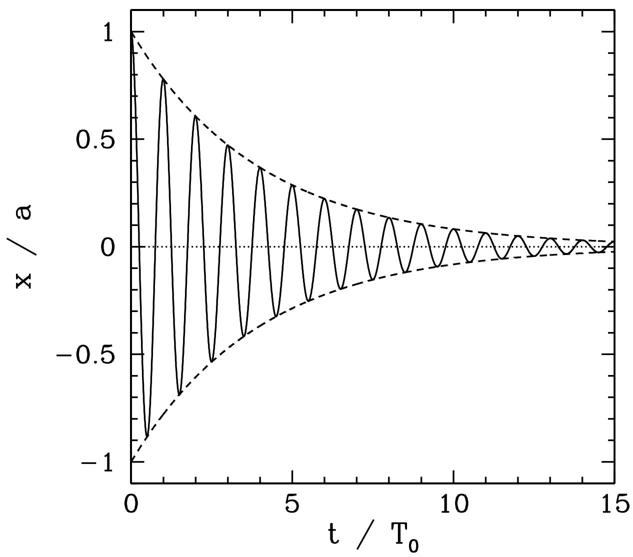


Figure 3.1: *Damped harmonic oscillation.*

oscillate at all, and any motion simply decays away exponentially in time (see Exercise 3).

Note that, although the angular frequency, ω_1, and decay rate, $\nu/2$, of the damped harmonic oscillation specified in Equation (3.11) are determined by the constants appearing in the damped harmonic oscillator equation, (3.2), the initial amplitude, a, and the phase angle, ϕ, of the oscillation are determined by the *initial conditions*. In fact, if $x(0) = x_0$ and $\dot{x}(0) = v_0$ then it follows from Equation (3.11) that

$$x_0 = a\cos\phi, \tag{3.12}$$

$$v_0 = -\frac{\nu}{2}\,a\cos\phi + \omega_1\,a\sin\phi, \tag{3.13}$$

giving

$$a = \left[x_0^2 + \frac{(v_0 + \nu\,x_0/2)^2}{\omega_1^2}\right]^{1/2}, \tag{3.14}$$

$$\phi = \tan^{-1}\left(\frac{v_0 + \nu\,x_0/2}{\omega_1\,x_0}\right). \tag{3.15}$$

Note, further, that the damped harmonic oscillator equation is a *linear* differential equation: *i.e.*, if $x(t)$ is a solution then so is $a\,x(t)$, where a is an arbitrary constant. It follows that the solutions of this equation are *superposable*, so that if $x_1(t)$ and $x_2(t)$ are two solutions corresponding to different initial conditions then $a\,x_1(t)+b\,x_2(t)$ is a third solution, where a and b are arbitrary constants.

Multiplying the damped harmonic oscillator equation (3.2) by $\dot{x}$, we obtain

$$\dot{x}\,\ddot{x}+\nu\,\dot{x}^{\,2}+\omega_0^2\,\dot{x}\,x=0, \tag{3.16}$$

which can be rearranged to give

$$\frac{dE}{dt}=-m\,\nu\,\dot{x}^{\,2}, \tag{3.17}$$

where

$$E=\frac{1}{2}\,m\,\dot{x}^{\,2}+\frac{1}{2}\,k\,x^2 \tag{3.18}$$

is the total energy of the system: *i.e.*, the sum of the kinetic and potential energies. Clearly, since the right-hand side of (3.17) cannot be positive, and is only zero when the system is stationary, the total energy is *not* a conserved quantity, but instead decays monotonically in time due to the presence of damping. Now, the net rate at which the force (3.1) does work on the mass is

$$P=f\,\dot{x}=-k\,\dot{x}\,x-m\,\nu\,\dot{x}^2. \tag{3.19}$$

Note that the spring force (*i.e.*, the first term on the right-hand side) does negative work on the mass (*i.e.*, it reduces the system kinetic energy) when $\dot{x}$ and x are of the same sign, and does positive work when they are of the opposite sign. On average, the spring force does no net work on the mass during an oscillation cycle. The damping force, on the other hand, (*i.e.*, the second term on the right-hand side) always does negative work on the mass, and, therefore, always acts to reduce the system kinetic energy.

3.2 Quality Factor

The energy loss rate of a weakly damped (*i.e.*, $\nu \ll 2\,\omega_0$) harmonic oscillator is conveniently characterized in terms of a parameter, Q_f, which is known as the *quality factor*. This quantity is defined to be 2π times the energy stored in the oscillator, divided by the energy lost in a single oscillation period. If the oscillator is weakly damped then the energy lost per period is relatively small, and Q_f is therefore much larger than unity. Roughly speaking, Q_f is the number of oscillations that the oscillator typically completes, after being set in motion, before its amplitude decays to a negligible value. For instance, the quality factor for the damped oscillation shown in Figure 3.1 is 12.6. Let us find an expression for Q_f.

As we have seen, the motion of a weakly damped harmonic oscillator is specified by [see Equation (3.11)]

$$x = a\,e^{-\nu\,t/2}\cos(\omega_1\,t - \phi), \tag{3.20}$$

It follows that

$$\dot{x} = -\frac{a\,\nu}{2}\,e^{-\nu\,t/2}\cos(\omega_1\,t-\phi) - a\,\omega_1\,e^{-\nu\,t/2}\sin(\omega_1\,t-\phi). \tag{3.21}$$

Thus, making use of Equation (3.17), the energy lost during a single oscillation period is

$$\Delta E = -\int_{\phi/\omega_1}^{(2\pi+\phi)/\omega_1}\frac{dE}{dt}\,dt \tag{3.22}$$

$$= m\,\nu\,a^2\int_{\phi/\omega_1}^{(2\pi+\phi)/\omega_1} e^{-\nu\,t}\left[\frac{\nu}{2}\cos(\omega_1\,t-\phi)+\omega_1\,\sin(\omega_1\,t-\phi)\right]^2 dt.$$

In the weakly damped limit, $\nu \ll 2\,\omega_0$, the exponential factor is approximately unity in the interval $t = \phi/\omega_1$ to $(2\pi+\phi)/\omega_1$, so that

$$\Delta E \simeq \frac{m\,\nu\,a^2}{\omega_1}\int_0^{2\pi}\left(\frac{\nu^2}{4}\cos^2\theta + \nu\,\omega_1\,\cos\theta\,\sin\theta + \omega_1^2\,\sin^2\theta\right)d\theta, \tag{3.23}$$

where $\theta = \omega_1\,t - \phi$. Thus,

$$\Delta E \simeq \frac{\pi\,m\,\nu\,a^2}{\omega_1}\,(\nu^2/4+\omega_1^2) = \pi\,m\,\omega_0^2\,a^2\left(\frac{\nu}{\omega_1}\right), \tag{3.24}$$

since, as is easily demonstrated,

$$\int_0^{2\pi}\cos^2\theta\,d\theta = \int_0^{2\pi}\sin^2\theta\,d\theta = \pi, \tag{3.25}$$

$$\int_0^{2\pi}\cos\theta\,\sin\theta\,d\theta = 0. \tag{3.26}$$

Now, the energy stored in the oscillator (at $t=0$) is [*cf.*, Equation (2.15)]

$$E = \frac{1}{2}\,m\,\omega_0^2\,a^2. \tag{3.27}$$

Hence, we obtain

$$Q_f = 2\pi\,\frac{E}{\Delta E} = \frac{\omega_1}{\nu} \simeq \frac{\omega_0}{\nu}. \tag{3.28}$$

3.3 LCR Circuit

Consider an electrical circuit consisting of an inductor, of inductance L, connected in series with a capacitor, of capacitance C, and a resistor, of resistance R. See Figure 3.2. Such a circuit is known as an LCR circuit, for obvious reasons. Suppose that $I(t)$ is the instantaneous current flowing around the circuit. As we saw in Section 2.3, the potential differences across the inductor and the capacitor are $L\,\dot{I}$ and Q/C, respectively. Here, Q is the charge on the capacitor's positive plate, and $I = \dot{Q}$. Moreover, from *Ohm's law*, the potential difference across the resistor is $V = I\,R$. Now, *Kichhoff's second circuital law* states that the sum of the potential differences across the various components of a closed circuit loop is zero. It follows that

$$L\,\dot{I} + R\,I + Q/C = 0. \tag{3.29}$$

Dividing by L, and differentiating with respect to time, we obtain

$$\ddot{I} + \nu\,\dot{I} + \omega_0^2\,I = 0, \tag{3.30}$$

where

$$\omega_0 = \frac{1}{\sqrt{L\,C}}, \tag{3.31}$$

$$\nu = \frac{R}{L}. \tag{3.32}$$

Comparison with Equation (3.2) reveals that (3.30) is a damped harmonic oscillator equation. Thus, provided that the resistance is not too high (*i.e.*, provided that $\nu < \omega_0/2$, which is equivalent to $R < 2\sqrt{L/C}$), the current in the circuit executes damped harmonic oscillations of the form [*cf.*, Equation (3.11)]

$$I(t) = I_0\,e^{-\nu\,t/2}\,\cos(\omega_1\,t - \phi), \tag{3.33}$$

where I_0 and ϕ are constants, and $\omega_1 = \sqrt{\omega_0^2 - \nu^2/4}$. We conclude that when a small amount of resistance is introduced into an LC circuit the characteristic oscillations in the current damp away exponentially at a rate proportional to the resistance.

Multiplying Equation (3.29) by I, and making use of the fact that $I = \dot{Q}$, we obtain

$$L\,\dot{I}\,I + R\,I^2 + \dot{Q}\,Q/C = 0, \tag{3.34}$$

which can be rearranged to give

$$\frac{dE}{dt} = -R\,I^2, \tag{3.35}$$

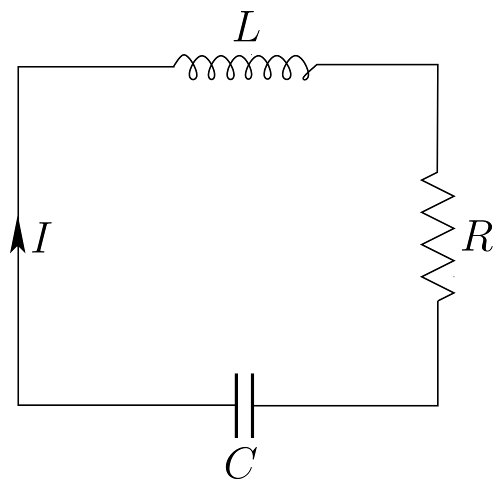


Figure 3.2: *An* LCR *circuit.*

where

$$E = \frac{1}{2}\,L\,I^2 + \frac{1}{2}\,\frac{Q^2}{C}. \tag{3.36}$$

Clearly, E is the circuit energy: *i.e.*, the sum of the energies stored in the inductor and the capacitor. Moreover, according to Equation (3.35), the circuit energy decays in time due to the power $R\,I^2$ dissipated via *Joule heating* in the resistor. Note that the dissipated power is always positive: *i.e.*, the circuit never gains energy from the resistor.

Finally, a comparison of Equations (3.28), (3.31), and (3.32) reveals that the quality factor of an LCR circuit is

$$Q_f = \frac{\sqrt{L/C}}{R}. \tag{3.37}$$

3.4 Driven Damped Harmonic Oscillation

We saw earlier, in Section 3.1, that when a damped mechanical oscillator is set into motion the oscillations eventually die away due to frictional energy losses. In fact, the only way of maintaining the motion of a damped oscillator is to continually feed energy into the system in such a manner as to offset the frictional losses. A steady-state (*i.e.*, constant amplitude) oscillation of this type is called *driven damped harmonic oscillation*. Consider a modified version of the mass-spring system investigated in Section 3.1 in

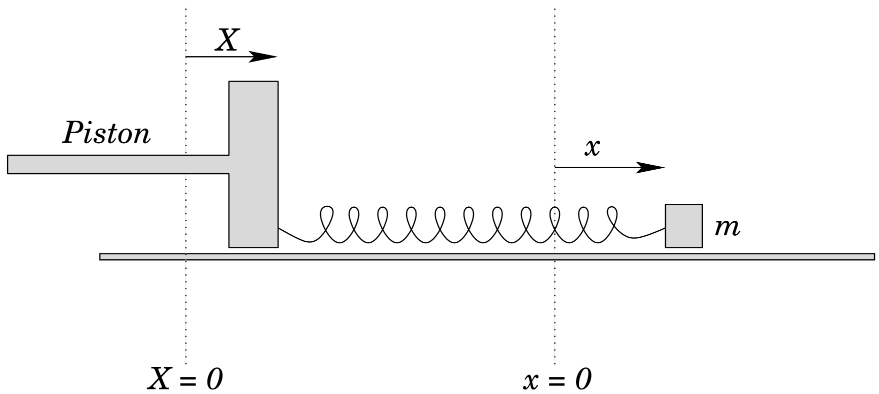


Figure 3.3: *A driven oscillatory system*

which one end of the spring is attached to the mass, and the other to a moving piston. See Figure 3.3. Let $x(t)$ be the horizontal displacement of the mass, and $X(t)$ the horizontal displacement of the piston. The extension of the spring is thus $x(t) - X(t)$, assuming that the spring is unstretched when $x = X = 0$. Thus, the horizontal force acting on the mass can be written [*cf.*, Equation (3.1)]

$$f = -k\,(x - X) - m\,\nu\,\dot{x}. \tag{3.38}$$

The equation of motion of the system then becomes [*cf.*, Equation (3.2)]

$$\ddot{x} + \nu\,\dot{x} + \omega_0^2\,x = \omega_0^2\,X, \tag{3.39}$$

where $\nu > 0$ is the damping constant, and $\omega_0 > 0$ the undamped oscillation frequency. Suppose, finally, that the piston executes *simple harmonic oscillation* of angular frequency $\omega > 0$ and amplitude $X_0 > 0$, so that the time evolution equation of the system takes the form

$$\ddot{x} + \nu\,\dot{x} + \omega_0^2\,x = \omega_0^2\,X_0\,\cos(\omega\,t). \tag{3.40}$$

We shall refer to the above as the *driven damped harmonic oscillator equation*.

Now, we would generally expect the periodically driven oscillator shown in Figure 3.3 to eventually settle down to a *steady* (*i.e.*, constant amplitude) pattern of oscillation, with the *same frequency* as the piston, in which the frictional energy loss per cycle is exactly matched by the work done by the piston per cycle (see Exercise 7). This suggests that we should search for a solution to Equation (3.40) of the form

$$x(t) = x_0\,\cos(\omega\,t - \varphi). \tag{3.41}$$

Here, $x_0 > 0$ is the *amplitude* of the driven oscillation, whereas φ is the *phase lag* of this oscillation (with respect to the phase of the piston oscillation). Now, since

$$\dot{x} = -\omega\, x_0 \sin(\omega\, t - \varphi), \tag{3.42}$$

$$\ddot{x} = -\omega^2\, x_0 \cos(\omega\, t - \varphi), \tag{3.43}$$

Equation (3.40) becomes

$$(\omega_0^2 - \omega^2)\, x_0 \cos(\omega\, t - \varphi) - \nu\, \omega\, x_0 \sin(\omega\, t - \varphi) = \omega_0^2\, X_0 \cos(\omega\, t). \tag{3.44}$$

However, $\cos(\omega\, t - \varphi) \equiv \cos(\omega\, t)\, \cos\varphi + \sin(\omega\, t)\, \sin\varphi$ and $\sin(\omega\, t - \varphi) \equiv \sin(\omega\, t)\, \cos\varphi - \cos(\omega\, t)\, \sin\varphi$, so we obtain

$$\begin{aligned} &\left[x_0\, (\omega_0^2 - \omega^2)\, \cos\varphi + x_0 \nu\, \omega\, \sin\varphi - \omega_0^2\, X_0\right] \cos(\omega\, t) \\ &\quad + x_0 \left[(\omega_0^2 - \omega^2)\, \sin\varphi - \nu\, \omega\, \cos\varphi\right] \sin(\omega\, t) = 0. \end{aligned} \tag{3.45}$$

Now, the only way in which the above equation can be satisfied at all times is if the coefficients of $\cos(\omega\, t)$ and $\sin(\omega\, t)$ separately equate to zero. In other words,

$$x_0\, (\omega_0^2 - \omega^2)\, \cos\varphi + x_0 \nu\, \omega\, \sin\varphi - \omega_0^2\, X_0 = 0, \tag{3.46}$$

$$(\omega_0^2 - \omega^2)\, \sin\varphi - \nu\, \omega\, \cos\varphi = 0. \tag{3.47}$$

These two expressions can be combined to give

$$x_0 = \frac{\omega_0^2\, X_0}{\left[(\omega_0^2 - \omega^2)^2 + \nu^2\, \omega^2\right]^{1/2}}, \tag{3.48}$$

$$\varphi = \tan^{-1}\left(\frac{\nu\, \omega}{\omega_0^2 - \omega^2}\right). \tag{3.49}$$

Let us investigate the dependence of the amplitude, x_0, and phase lag, φ, of the driven oscillation on the driving frequency, ω. This is most easily done graphically. Figure 3.4 shows x_0/X_0 and φ plotted as functions of ω for various different values of ν/ω_0. In fact, $\nu/\omega_0 = 1/Q_f = 1$, $1/2$, $1/4$, $1/8$, and $1/16$ correspond to the solid, dotted, short-dashed, long-dashed, and dot-dashed curves, respectively. It can be seen that as the amount of damping in the system is decreased the amplitude of the response becomes progressively more peaked at the natural frequency of oscillation of the system, ω_0. This effect is known as *resonance*, and ω_0 is termed the *resonant*

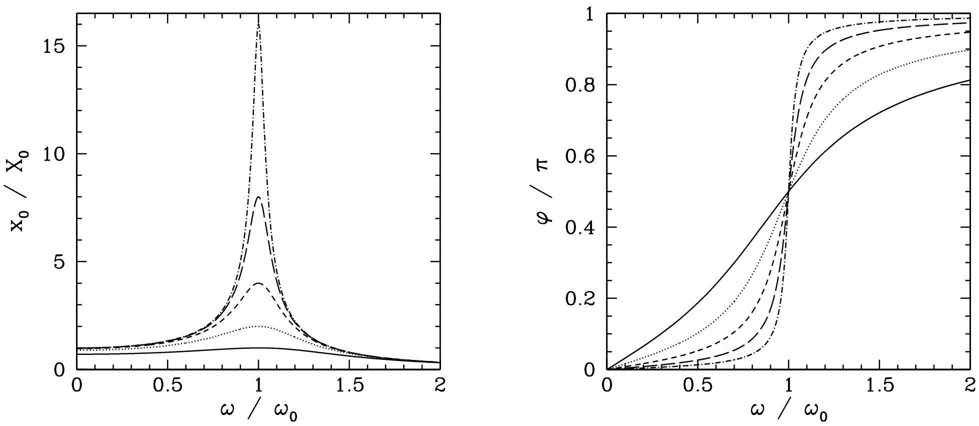


Figure 3.4: *Driven harmonic motion.*

frequency. Thus, a weakly damped oscillator (*i.e.*, $\nu \ll \omega_0$) can be driven to large amplitude by the application of a relatively small amplitude external driving force which oscillates at a frequency close to the resonant frequency. Note that the response of the oscillator is *in phase* (*i.e.*, $\varphi \simeq 0$) with the external drive for driving frequencies well below the resonant frequency, is *in phase quadrature* (*i.e.*, $\varphi = \pi/2$) at the resonant frequency, and is *in anti-phase* (*i.e.*, $\varphi \simeq \pi$) for frequencies well above the resonant frequency.

According to Equations (3.28) and (3.48),

$$\frac{x_0(\omega = \omega_0)}{X_0} = \frac{\omega_0}{\nu} = Q_f. \tag{3.50}$$

In other words, when the driving frequency matches the resonant frequency the ratio of the amplitude of the driven oscillation to that of the piston oscillation is the quality factor, Q_f. Hence, Q_f can be regarded as the *resonant amplification factor* of the oscillator. Equations (3.48) and (3.49) imply that, for a weakly damped oscillator (*i.e.*, $\nu \ll \omega_0$) which is close to resonance [*i.e.*, $|\omega - \omega_0| \sim \nu \ll \omega_0$],

$$\frac{x_0(\omega)}{x_0(\omega = \omega_0)} \simeq \sin\varphi \simeq \frac{\nu}{[4\,(\omega_0 - \omega)^2 + \nu^2]^{1/2}}. \tag{3.51}$$

Hence, the width of the resonance peak (in angular frequency) is $\Delta\omega = \nu$, where the edges of peak are defined as the points at which the driven amplitude is reduced to $1/\sqrt{2}$ of its maximum value: *i.e.*, $\omega = \omega_0 \pm \nu/2$. Note that the phase lag at the low and high frequency edges of the peak are

$\pi/4$ and $3\pi/4$, respectively. Furthermore, the fractional width of the peak is

$$\frac{\Delta\omega}{\omega_0} = \frac{\nu}{\omega_0} = \frac{1}{Q_f}. \tag{3.52}$$

We conclude that the height and width of the resonance peak of a weakly damped ($Q_f \gg 1$) harmonic oscillator scale as Q_f and Q_f^{-1}, respectively. Thus, the area under the resonance peak stays approximately constant as Q_f varies.

Now, the force exerted on the system by the piston is

$$F(t) = k\,X_0\,\cos(\omega\,t). \tag{3.53}$$

Hence, the instantaneous power absorption from the piston becomes

$$\begin{aligned} P(t) &= F(t)\,\dot{x}(t) \\ &= -k\,X_0\,x_0\,\omega\,\cos(\omega\,t)\,\sin(\omega\,t-\varphi) \\ &= -k\,X_0\,x_0\,\omega\left[\cos(\omega\,t)\,\sin(\omega\,t)\,\cos\varphi - \cos^2(\omega\,t)\,\sin\varphi\right]. \end{aligned} \tag{3.54}$$

Thus, the average power absorption during an oscillation cycle is

$$\langle P\rangle = \frac{1}{2}\,k\,X_0\,x_0\,\omega\,\sin\varphi, \tag{3.55}$$

since $\langle\cos(\omega\,t)\,\sin(\omega\,t)\rangle = 0$ and $\langle\cos^2(\omega\,t)\rangle = 1/2$. Of course, given that the amplitude of the driven oscillation neither grows nor decays, the average power absorption from the piston during an oscillation cycle must be equal to the average power dissipation due to friction (see Exercise 7). Making use of Equations (3.50) and (3.51), the mean power absorption when the driving frequency is close to the resonant frequency is

$$\langle P\rangle \simeq \frac{1}{2}\,\omega_0\,k\,X_0^2\,Q_f\left[\frac{\nu^2}{4\,(\omega_0-\omega)^2+\nu^2}\right]. \tag{3.56}$$

Thus, the maximum power absorption occurs at the resonance (*i.e.*, $\omega = \omega_0$), and the absorption is reduced to half of this maximum value at the edges of the resonance (*i.e.*, $\omega = \omega_0 \pm \nu/2$). Furthermore, the peak power absorption is proportional to the quality factor, Q_f, which means that it is inversely proportional to the damping constant, ν.

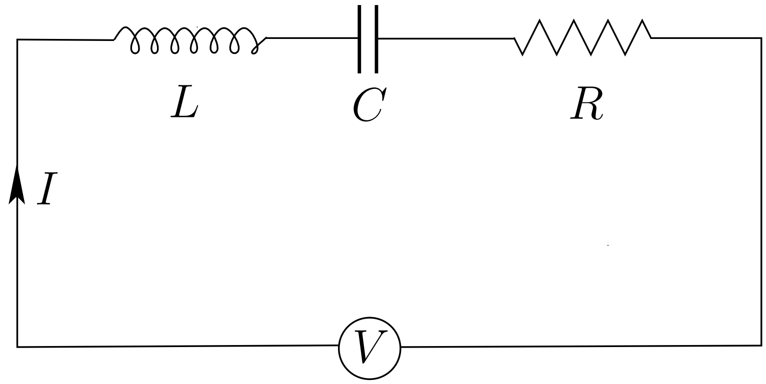


Figure 3.5: *A driven* LCR *circuit.*

3.5 Driven LCR Circuit

Consider an LCR circuit consisting of an inductor, L, a capacitor, C, and a resistor, R, connected in series with an emf of voltage $V(t)$. See Figure 3.5. Let $I(t)$ be the instantaneous current flowing around the circuit. Now, according to *Kichhoff's second circuital law*, the sum of the potential drops across the various components of a closed circuit loop is equal to zero. Thus, since the potential drop across an emf is *minus* the associated voltage, we obtain [*cf.*, Equation (3.29)]

$$L\,\dot{I} + R\,I + Q/C = V, \tag{3.57}$$

where $\dot{Q} = I$. Suppose that the emf is such that its voltage oscillates *sinusoidally* at the angular frequency $\omega > 0$, with the peak value $V_0 > 0$, so that

$$V(t) = V_0\,\sin(\omega\,t). \tag{3.58}$$

Dividing Equation (3.57) by L, and differentiating with respect to time, we obtain [*cf.*, Equation (3.30)]

$$\ddot{I} + \nu\,\dot{I} + \omega_0^2\,I = \frac{\omega\,V_0}{L}\,\cos(\omega\,t), \tag{3.59}$$

where $\omega_0 = 1/\sqrt{L\,C}$ and $\nu = R/L$. Comparison with Equation (3.40) reveals that this is a driven damped harmonic oscillator equation. It follows, by comparison with the analysis contained in the previous section, that the current driven in the circuit by the oscillating emf is of the form

$$I(t) = I_0\,\cos(\omega\,t - \varphi), \tag{3.60}$$

where

$$
\begin{aligned}
I_0 &= \frac{\omega\, V_0/L}{\left[(\omega_0^2-\omega^2)^2+\nu^2\,\omega^2\right]^{1/2}}, && (3.61)\\
\varphi &= \tan^{-1}\left(\frac{\nu\,\omega}{\omega_0^2-\omega^2}\right). && (3.62)
\end{aligned}
$$

In the immediate vicinity of the resonance (*i.e.*, $|\omega-\omega_0|\sim\nu\ll\omega_0$) these expression simplify to

$$
\begin{aligned}
I_0 &\simeq \frac{V_0}{R}\,\frac{\nu}{[4\,(\omega_0-\omega)^2+\nu^2]^{1/2}}, && (3.63)\\
\sin\varphi &\simeq \frac{\nu}{[4\,(\omega-\omega_0)^2+\nu^2]^{1/2}}. && (3.64)
\end{aligned}
$$

Now, the circuit's mean power absorption from the emf is written

$$
\begin{aligned}
\langle P\rangle &= \langle I(t)\,V(t)\rangle = I_0\,V_0\,\langle\cos(\omega\,t-\varphi)\,\sin(\omega\,t)\rangle\\
&= \frac{1}{2}\,I_0\,V_0\,\sin\varphi, && (3.65)
\end{aligned}
$$

so that

$$
\langle P\rangle \simeq \frac{1}{2}\,\frac{V_0^2}{R}\left[\frac{\nu^2}{4\,(\omega_0-\omega)^2+\nu^2}\right] \tag{3.66}
$$

close to the resonance. It follows that the peak power absorption, $(1/2)\,V_0^2/R$, takes place when the emf oscillates at the resonant frequency, ω_0. Moreover, the power absorption falls to half of this peak value at the edges of the resonant peak: *i.e.*, $\omega=\omega_0\pm\nu$.

LCR circuits are often employed as analogue *radio tuners*. In this application, the emf represents the analogue signal picked-up by a radio antenna. It is clear, from the above analysis, that the circuit only has a strong response (*i.e.*, it only absorbs significant energy) when the signal oscillates in the angular frequency range $\omega_0\pm\nu$, which corresponds to $1/\sqrt{L\,C}\pm R/L$. Thus, if the values of L, C, and R are properly chosen then the circuit can be made to strongly absorb the signal from a particular radio station, which has a given carrier frequency and bandwidth, whilst essentially ignoring the signals from other stations with different carrier frequencies. In practice, the values of L and R are fixed, whilst the value of C is varied (by turning a knob which adjusts the degree of overlap between two sets of parallel semi-circular conducting plates) until the signal from the desired radio station is found.

3.6 Transient Oscillator Response

The time evolution of the driven mechanical oscillator discussed in Section 3.4 is governed by the *driven damped harmonic oscillator equation*,

$$\ddot{x} + \nu\,\dot{x} + \omega_0^2\,x = \omega_0^2\,X_0\,\cos(\omega\,t). \tag{3.67}$$

Recall that the steady-state (*i.e.*, constant amplitude) solution to this equation which we found earlier takes the form

$$x_{ta}(t) = x_0\,\cos(\omega\,t - \varphi), \tag{3.68}$$

where

$$x_0 = \frac{\omega_0^2\,X_0}{\left[(\omega_0^2-\omega^2)^2+\nu^2\,\omega^2\right]^{1/2}}, \tag{3.69}$$

$$\varphi = \tan^{-1}\left(\frac{\nu\,\omega}{\omega_0^2-\omega^2}\right). \tag{3.70}$$

Now, Equation (3.67) is a *second-order* ordinary differential equation, which means that its general solution should contain *two* arbitrary constants. Note, however, that (3.68) contains *no* arbitrary constants. It follows that (3.68) cannot be the most general solution to the driven damped harmonic oscillator equation, (3.67). However, it is fairly easy to see that if we add any solution of the *undriven damped harmonic oscillator equation*,

$$\ddot{x} + \nu\,\dot{x} + \omega_0^2\,x = 0, \tag{3.71}$$

to (3.68) then the result will still be a solution to Equation (3.67). Now, from Section 3.1, the most general solution to the above equation can be written

$$x_{tr}(t) = A\,e^{-\nu\,t/2}\,\cos(\omega_1\,t) + B\,e^{-\nu\,t/2}\,\sin(\omega_1\,t), \tag{3.72}$$

where $\omega_1 = (\omega_0^2 - \nu^2/4)^{1/2}$, and A and B are arbitrary constants. [In terms of the standard solution (3.11), $A = a\,\cos\phi$ and $B = a\,\sin\phi$.] Thus, a more general solution to (3.67) is

$$\begin{aligned} x(t) &= x_{ta}(t) + x_{tr}(t) \qquad (3.73)\\ &= x_0\,\cos(\omega\,t-\varphi) + A\,e^{-\nu\,t/2}\,\cos(\omega_1\,t) + B\,e^{-\nu\,t/2}\,\sin(\omega_1\,t). \end{aligned}$$

In fact, since the above solution contains *two* arbitrary constants, we can be sure that it is the *most general* solution. Of course, the constants A and B are determined by the *initial conditions*. It is, thus, clear that the most general solution to the driven damped harmonic oscillator equation (3.67) consists of *two* parts. First, the solution (3.68), which oscillates at the *driving frequency* ω with a *constant* amplitude, and which is *independent* of the initial conditions. Second, the solution (3.72), which oscillates at the *natural frequency* ω_1 with an amplitude which *decays exponentially* in time, and which *depends* on the initial conditions. The former is termed the *time asymptotic solution*, since if we wait long enough then it becomes dominant. The latter is called the *transient solution*, since if we wait long enough then it decays away.

Suppose, for the sake of argument, that the system is initially in its equilibrium state: *i.e.*, $x(0) = \dot{x}(0) = 0$. It follows from (3.73) that

$$x(0) = x_0 \cos\varphi + A = 0, \tag{3.74}$$

$$\dot{x}(0) = x_0\,\omega\,\sin\varphi - \frac{\nu}{2}\,A + B\,\omega_1 = 0. \tag{3.75}$$

These equations can be solved to give

$$A = -x_0 \cos\varphi, \tag{3.76}$$

$$B = -x_0\left[\frac{\omega\,\sin\varphi + (\nu/2)\,\cos\varphi}{\omega_1}\right]. \tag{3.77}$$

Now, according to the analysis in Section 3.4, for driving frequencies close to the resonant frequency (*i.e.*, $|\omega - \omega_0| \sim \nu$), we can write

$$x_0 \simeq \frac{X_0\,\omega_0}{[4\,(\omega_0-\omega)^2+\nu^2]^{1/2}}, \tag{3.78}$$

$$\sin\varphi \simeq \frac{\nu}{[4\,(\omega_0-\omega)^2+\nu^2]^{1/2}}, \tag{3.79}$$

$$\cos\varphi \simeq \frac{2\,(\omega_0-\omega)}{[4\,(\omega_0-\omega)^2+\nu^2]^{1/2}}. \tag{3.80}$$

Hence, in this case, the solution (3.73), combined with (3.76)–(3.80), reduces to

$$\begin{aligned} x(t) \simeq\; & X_0\left[\frac{2\,\omega_0\,(\omega_0-\omega)}{4\,(\omega_0-\omega)^2+\nu^2}\right]\left[\cos(\omega\,t) - e^{-\nu\,t/2}\cos(\omega_0\,t)\right] \\ & + X_0\left[\frac{\omega_0\,\nu}{4\,(\omega_0-\omega)^2+\nu^2}\right]\left[\sin(\omega\,t) - e^{-\nu\,t/2}\sin(\omega_0\,t)\right]. \end{aligned} \tag{3.81}$$

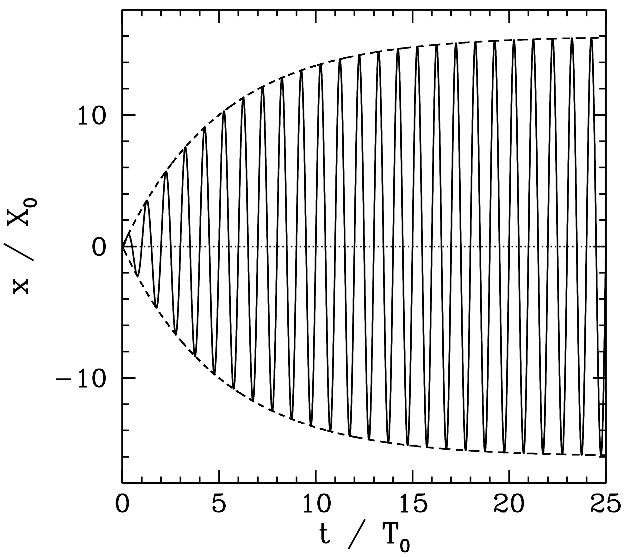


Figure 3.6: *Resonant response of a driven damped harmonic oscillator.*

There are a number of interesting cases which are worthy of discussion. Consider, first, the situation in which the driving frequency is equal to the resonant frequency: *i.e.*, $\omega = \omega_0$. In this case, Equation (3.81) reduces to

$$x(t) = X_0\, Q_f \left(1 - e^{-\nu\, t/2}\right) \sin(\omega_0\, t), \tag{3.82}$$

since $Q_f = \omega_0/\nu$. Thus, the driven response oscillates at the resonant frequency, ω_0, since both the time asymptotic and transient solutions oscillate at this frequency. However, the amplitude of the oscillation grows monotonically as $a(t) = X_0\, Q_f \left(1 - e^{-\nu\, t/2}\right)$, and so takes a time of order ν^{-1} to attain its final value $X_0\, Q_f$, which is, of course, larger that the driving amplitude by the resonant amplification factor (or quality factor), Q_f. This behavior is illustrated in Figure 3.6. [Here, $T_0 = 2\pi/\omega_0$ and $Q_f = \omega_0/\nu = 16$. The solid curve shows $x(t)/X_0$ and the dashed curves show $\pm a(t)/X_0$.]

Consider, now, the situation in which there is no damping, so that $\nu = 0$. In this case, Equation (3.81) yields

$$x(t) = X_0 \left(\frac{\omega_0}{\omega_0 - \omega}\right) \sin[(\omega_0 - \omega)\, t/2]\, \sin[(\omega_0 + \omega)\, t/2], \tag{3.83}$$

where use has been made of the trigonometry identity $\cos a - \cos b \equiv -2\, \sin[(a+b)/2]\, \sin[(a-b)/2]$. It can be seen that the driven response

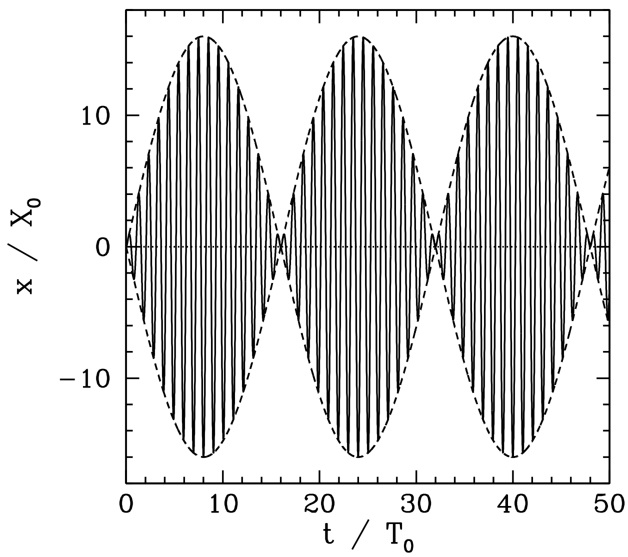


Figure 3.7: *Off-resonant response of a driven undamped harmonic oscillator.*

oscillates relatively rapidly at the "sum frequency" $(\omega_0 + \omega)/2$ with an amplitude $a(t) = X_0\,[\omega_0/(\omega_0 - \omega)]\,\sin[(\omega_0 - \omega)/t]$ which modulates relatively slowly at the "difference frequency" $(\omega_0 - \omega)/2$. (Recall, that we are assuming that ω is close to ω_0.) This behavior is illustrated in Figure 3.7. [Here, $T_0 = 2\pi/\omega_0$ and $\omega_0 - \omega = \omega_0/16$. The solid curve shows $x(t)/X_0$ and the dashed curves show $\pm a(t)/X_0$.] The amplitude modulations shown in Figure 3.7 are called *beats*, and are produced whenever two sinusoidal oscillations of similar amplitude, and slightly different frequency, are superposed. In this case, the two oscillations are the time asymptotic solution, which oscillates at the driving frequency, ω, and the transient solution, which oscillates at the resonant frequency, ω_0. The beats modulate at the difference frequency, $(\omega_0 - \omega)/2$. In the limit $\omega \to \omega_0$, Equation (3.83) yields

$$x(t) = \frac{X_0}{2}\,\omega_0\,t\,\sin(\omega_0\,t), \tag{3.84}$$

since $\sin x \simeq x$ when $|x| \ll 1$. Thus, the resonant response of a driven undamped oscillator is an oscillation at the resonant frequency whose amplitude, $a(t) = (X_0/2)\,\omega_0\,t$, increases *linearly* in time. In this case, the period of the beats has effectively become infinite.

Finally, Figure 3.8 illustrates the non-resonant response of a driven damped harmonic oscillator, obtained from Equation (3.81). [Here, $T_0 = 2\pi/\omega_0$,

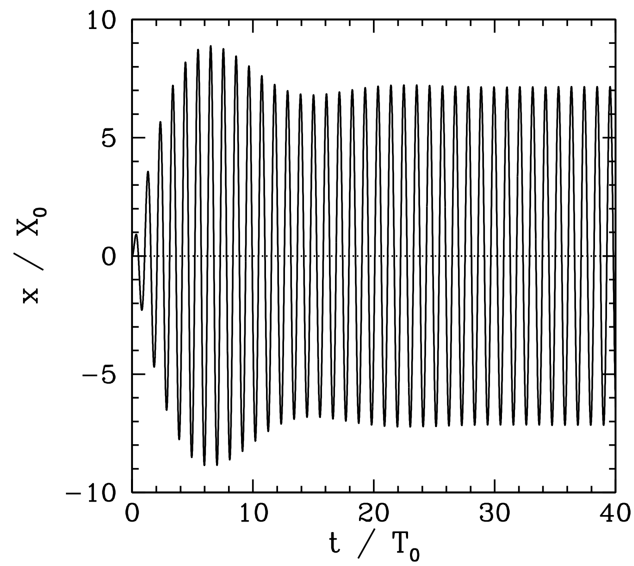


Figure 3.8: *Off-resonant response of a driven damped harmonic oscillator.*

$\omega_0 - \omega = \omega_0/16$, and $\nu = \omega_0/16$.] It can be seen that the driven response grows, showing some initial evidence of beat modulation, but eventually settles down to a steady pattern of oscillation. This behavior occurs because the transient solution, which is needed to produce beats, initially grows, but then damps away, leaving behind the constant amplitude time asymptotic solution.

3.7 Exercises

1. Show that the ratio of two successive maxima in the displacement of a damped harmonic oscillator is constant.

2. If the amplitude of a damped harmonic oscillator decreases to $1/e$ of its initial value after $n \gg 1$ periods show that the ratio of the period of oscillation to the period of the oscillation with no damping is

$$\left(1 + \frac{1}{4\,\pi^2\,n^2}\right)^{1/2} \simeq 1 + \frac{1}{8\,\pi^2\,n^2}.$$

3. Many oscillatory systems are subject to damping effects which are not exactly analogous to the frictional damping considered in Section 3.1. Nevertheless, such systems typically exhibit an exponential decrease in their average stored

energy of the form $\langle E \rangle = E_0 \exp(-\nu t)$. It is possible to define an effective quality factor for such oscillators as $Q_f = \omega_0/\nu$, where ω_0 is the natural angular oscillation frequency. For example, when the note "middle C" on a piano is struck its oscillation energy decreases to one half of its initial value in about 1 second. The frequency of middle C is 256 Hz. What is the effective Q_f of the system?

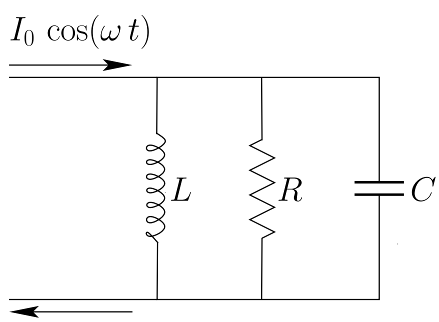
4. According to classical electromagnetic theory, an accelerated electron radiates energy at the rate $K e^2 a^2/c^3$, where $K = 6 \times 10^9 \, \mathrm{N\,m^2/C^2}$, e is the charge on an electron, a the instantaneous acceleration, and c the velocity of light. If an electron were oscillating in a straight-line with displacement $x(t) = A \sin(2\pi f t)$ how much energy would it radiate away during a single cycle? What is the effective Q_f of this oscillator? How many periods of oscillation would elapse before the energy of the oscillation was reduced to half of its initial value? Substituting a typical optical frequency (*i.e.*, for visible light) for f, give numerical estimates for the Q_f and half-life of the radiating system.

5. Demonstrate that in the limit $\nu \to 2\,\omega_0$ the solution to the damped harmonic oscillator equation becomes

$$x(t) = (x_0 + [v_0 + (\nu/2)\, x_0]\, t)\; e^{-\nu t/2},$$

where $x_0 = x(0)$ and $v_0 = \dot{x}(0)$.

6. What are the resonant angular frequency and quality factor of the circuit pictured below? What is the average power absorbed at resonance?



7. The power input $\langle P \rangle$ required to maintain a constant amplitude oscillation in a driven damped harmonic oscillator can be calculated by recognizing that this power is minus the average rate that work is done by the damping force, $-k\,\dot{x}$.

   (a) Using $x = x_0 \cos(\omega t - \varphi)$, show that the average rate that the damping force does work is $-k\,\omega^2\,x_0^2/2$.

   (b) Substitute the value of x_0 at an arbitrary driving frequency and, hence, obtain an expression for $\langle P \rangle$.

(c) Demonstrate that this expression yields (3.56) in the limit that the driving frequency is close to the resonant frequency.

8. A generator of emf $V(t) = V_0 \cos(\omega t)$ is connected in series with a resistance R, an inductor L, and a capacitor C. Let $I(t)$ be the current flowing in the circuit, and $Q(t)$ the charge on the capacitor. Suppose that $I = Q = 0$ at $t = 0$. Find $I(t)$ and $Q(t)$ for $t > 0$.

9. The equation $m\ddot{x} + k x = F_0 \sin(\omega t)$ governs the motion of an undamped harmonic oscillator driven by a sinusoidal force of angular frequency ω. Show that the steady-state solution is

$$x = \frac{F_0 \sin(\omega t)}{m(\omega_0^2 - \omega^2)},$$

where $\omega_0 = \sqrt{k/m}$. Sketch the behavior of x versus t for $\omega < \omega_0$ and $\omega > \omega_0$. Demonstrate that if $x = \dot{x} = 0$ at $t = 0$ then the general solution is

$$x = \frac{F_0}{m(\omega_0^2 - \omega^2)}\left[\sin(\omega t) - \frac{\omega}{\omega_0}\sin(\omega_0 t)\right].$$

Show, finally, that if ω is close to the resonant frequency ω_0 then

$$x \simeq \frac{F_0}{2\omega_0^2}\left[\sin(\omega_0 t) - \omega_0 t \cos(\omega_0 t)\right].$$

Sketch the behavior of x versus t.

4 Coupled Oscillations

4.1 Two Spring-Coupled Masses

Consider a mechanical system consisting of two identical masses m which are free to slide over a frictionless horizontal surface. Suppose that the masses are attached to one another, and to two immovable walls, by means of three identical light horizontal springs of spring constant k, as shown in Figure 4.1. The instantaneous state of the system is conveniently specified by the displacements of the left and right masses, $x_1(t)$ and $x_2(t)$, respectively. The extensions of the left, middle, and right springs are thus x_1, $x_2 - x_1$, and $-x_2$, respectively, assuming that $x_1 = x_2 = 0$ corresponds to the equilibrium configuration in which the springs are all unextended. The equations of motion of the two masses are thus

$$m\,\ddot{x}_1 = -k\,x_1 + k\,(x_2 - x_1), \tag{4.1}$$

$$m\,\ddot{x}_1 = -k\,(x_2 - x_1) + k\,(-x_2). \tag{4.2}$$

Here, we have made use of the fact that a mass attached to the left end of a spring of extension x and spring constant k experiences a horizontal force $+k\,x$, whereas a mass attached to the right end of the same spring experiences an equal and opposite force $-k\,x$.

Equations (4.1)–(4.2) can be rewritten in the form

$$\ddot{x}_1 = -2\,\omega_0^2\,x_1 + \omega_0^2\,x_2, \tag{4.3}$$

$$\ddot{x}_2 = \omega_0^2\,x_1 - 2\,\omega_0^2\,x_2, \tag{4.4}$$

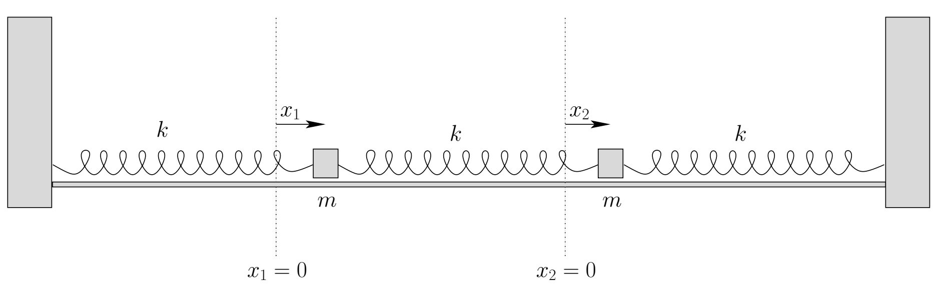


Figure 4.1: *Two degree of freedom mass-spring system.*

where $\omega_0 = \sqrt{k/m}$. Let us search for a solution in which the two masses oscillate *in phase* at the *same angular frequency*, ω. In other words,

$$x_1(t) = \hat{x}_1 \cos(\omega\,t - \phi), \tag{4.5}$$

$$x_2(t) = \hat{x}_2 \cos(\omega\,t - \phi), \tag{4.6}$$

where $\hat{x}_1$, $\hat{x}_2$, and ϕ are constants. Equations (4.3) and (4.4) yield

$$-\,\omega^2\,\hat{x}_1 \cos(\omega\,t - \phi) = \left(-2\,\omega_0^2\,\hat{x}_1 + \omega_0^2\,\hat{x}_2\right) \cos(\omega\,t - \phi), \tag{4.7}$$

$$-\omega^2\,\hat{x}_2 \cos(\omega\,t - \phi) = \left(\omega_0^2\,\hat{x}_1 - 2\,\omega_0^2\,\hat{x}_2\right) \cos(\omega\,t - \phi), \tag{4.8}$$

or

$$(\hat{\omega}^{\,2} - 2)\,\hat{x}_1 + \hat{x}_2 = 0, \tag{4.9}$$

$$\hat{x}_1 + (\hat{\omega}^{\,2} - 2)\,\hat{x}_2 = 0, \tag{4.10}$$

where $\hat{\omega} = \omega/\omega_0$. Note that by searching for a solution of the form (4.5)–(4.6) we have effectively converted the system of two coupled *linear differential equations* (4.3)–(4.4) into the much simpler system of two coupled *linear algebraic equations* (4.9)–(4.10). The latter equations have the trivial solutions $\hat{x}_1 = \hat{x}_2 = 0$, but also yield

$$\frac{\hat{x}_1}{\hat{x}_2} = -\frac{1}{(\hat{\omega}^{\,2} - 2)} = -(\hat{\omega}^{\,2} - 2). \tag{4.11}$$

Hence, the condition for a nontrivial solution is clearly

$$(\hat{\omega}^{\,2} - 2)\,(\hat{\omega}^{\,2} - 2) - 1 = 0. \tag{4.12}$$

In fact, if we write Equations (4.9)–(4.10) in the form of a homogenous (*i.e.*, with a null right-hand side) 2×2 matrix equation, so that

$$\begin{pmatrix} \hat{\omega}^{\,2} - 2 & 1 \\ 1 & \hat{\omega}^{\,2} - 2 \end{pmatrix} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{4.13}$$

then it is clear that the criterion (4.12) can also be obtained by setting the *determinant* of the associated 2×2 matrix to *zero*.

Equation (4.12) can be rewritten

$$\hat{\omega}^{\,4} - 4\,\hat{\omega}^2 + 3 = (\hat{\omega}^{\,2} - 1)\,(\hat{\omega}^{\,2} - 3) = 0. \tag{4.14}$$

It follows that

$$\hat{\omega} = 1 \text{ or } \sqrt{3}. \tag{4.15}$$

Here, we have neglected the two negative frequency roots of (4.14)—*i.e.*, $\hat{\omega} = -1$ and $\hat{\omega} = -\sqrt{3}$—since a negative frequency oscillation is equivalent to an oscillation with an equal and opposite positive frequency, and an equal and opposite phase: *i.e.*, $\cos(\omega\, t - \phi) \equiv \cos(-\omega\, t + \phi)$. It is thus apparent that the dynamical system pictured in Figure 4.1 has *two* unique frequencies of oscillation: *i.e.*, $\omega = \omega_0$ and $\omega = \sqrt{3}\,\omega_0$. These are called the *normal frequencies* of the system. Since the system possesses *two degrees of freedom* (*i.e.*, two independent coordinates are needed to specify its instantaneous configuration) it is not entirely surprising that it possesses *two* normal frequencies. In fact, it is a general rule that a dynamical system with N degrees of freedom possesses N normal frequencies.

The patterns of motion associated with the two normal frequencies can easily be deduced from Equation (4.11). Thus, for $\omega = \omega_0$ (*i.e.*, $\hat{\omega} = 1$), we get $\hat{x}_1 = \hat{x}_2$, so that

$$x_1(t) = \hat{\eta}_1 \cos(\omega_0\, t - \phi_1), \tag{4.16}$$

$$x_2(t) = \hat{\eta}_1 \cos(\omega_0\, t - \phi_1), \tag{4.17}$$

where $\hat{\eta}_1$ and ϕ_1 are constants. This first pattern of motion corresponds to the two masses executing simple harmonic oscillation with the *same amplitude and phase*. Note that such an oscillation does not stretch the middle spring. On the other hand, for $\omega = \sqrt{3}\,\omega_0$ (*i.e.*, $\hat{\omega} = \sqrt{3}$), we get $\hat{x}_1 = -\hat{x}_2$, so that

$$x_1(t) = \hat{\eta}_2 \cos\left(\sqrt{3}\,\omega_0\, t - \phi_2\right), \tag{4.18}$$

$$x_2(t) = -\hat{\eta}_2 \cos\left(\sqrt{3}\,\omega_0\, t - \phi_2\right), \tag{4.19}$$

where $\hat{\eta}_2$ and ϕ_2 are constants. This second pattern of motion corresponds to the two masses executing simple harmonic oscillation with the *same amplitude but in anti-phase*: *i.e.*, with a phase shift of π radians. Such oscillations do stretch the middle spring, implying that the restoring force associated with similar amplitude displacements is greater for the second pattern of motion than for the first. This accounts for the higher oscillation frequency in the second case. (The inertia is the same in both cases, so the oscillation frequency is proportional to the square root of the restoring force associated with similar amplitude displacements.) The two distinctive patterns of motion which we have found are called the *normal modes* *of oscillation* of the system. Incidentally, it is a general rule that a dynamical system possessing N degrees of freedom has N unique normal modes of oscillation.

Now, the most general motion of the system is a *linear combination* of the two normal modes. This immediately follows because Equations (4.1) and (4.2) are *linear equations*. [In other words, if $x_1(t)$ and $x_2(t)$ are solutions then so are $a\,x_1(t)$ and $a\,x_2(t)$, where a is an arbitrary constant.] Thus, we can write

$$x_1(t) = \hat{\eta}_1 \cos(\omega_0\,t - \phi_1) + \hat{\eta}_2 \cos\left(\sqrt{3}\,\omega_0\,t - \phi_2\right), \tag{4.20}$$

$$x_2(t) = \hat{\eta}_1 \cos(\omega_0\,t - \phi_1) - \hat{\eta}_2 \cos\left(\sqrt{3}\,\omega_0\,t - \phi_2\right). \tag{4.21}$$

Note that we can be sure that this represents the most general solution to Equations (4.1) and (4.2) because it contains *four* arbitrary constants: *i.e.*, $\hat{\eta}_1$, ϕ_1, $\hat{\eta}_2$, and ϕ_2. (In general, we expect the solution of a second-order ordinary differential equation to contain two arbitrary constants. It, thus, follows that the solution of a system of two coupled ordinary differential equations should contain four arbitrary constants.) Of course, these constants are determined by the *initial conditions*.

For instance, suppose that $x_1 = a$, $\dot{x}_1 = 0$, $x_2 = 0$, and $\dot{x}_2 = 0$ at $t = 0$. It follows, from (4.20) and (4.21), that

$$a = \hat{\eta}_1 \cos\phi_1 + \hat{\eta}_2 \cos\phi_2, \tag{4.22}$$

$$0 = \hat{\eta}_1 \sin\phi_1 + \sqrt{3}\,\hat{\eta}_2 \sin\phi_2, \tag{4.23}$$

$$0 = \hat{\eta}_1 \cos\phi_1 - \hat{\eta}_2 \cos\phi_2, \tag{4.24}$$

$$0 = \hat{\eta}_1 \sin\phi_1 - \sqrt{3}\,\hat{\eta}_2 \sin\phi_2, \tag{4.25}$$

which implies that $\phi_1 = \phi_2 = 0$ and $\hat{\eta}_1 = \hat{\eta}_2 = a/2$. Thus, the system evolves in time as

$$x_1(t) = a\,\cos(\omega_-\,t)\,\cos(\omega_+\,t), \tag{4.26}$$

$$x_2(t) = a\,\sin(\omega_-\,t)\,\sin(\omega_+\,t), \tag{4.27}$$

where $\omega_\pm = [(\sqrt{3} \pm 1)/2]\,\omega_0$, and use has been made of the trigonometric identities $\cos(a+b) \equiv 2\,\cos[(a+b)/2]\,\cos[(a-b)/2]$ and $\cos(a-b) \equiv -2\,\sin[(a+b)/2]\,\sin[(a-b)/2]$. This evolution is illustrated in Figure 4.2. [Here, $T_0 = 2\pi/\omega_0$. The solid curve corresponds to x_1, and the dashed curve to x_2.]

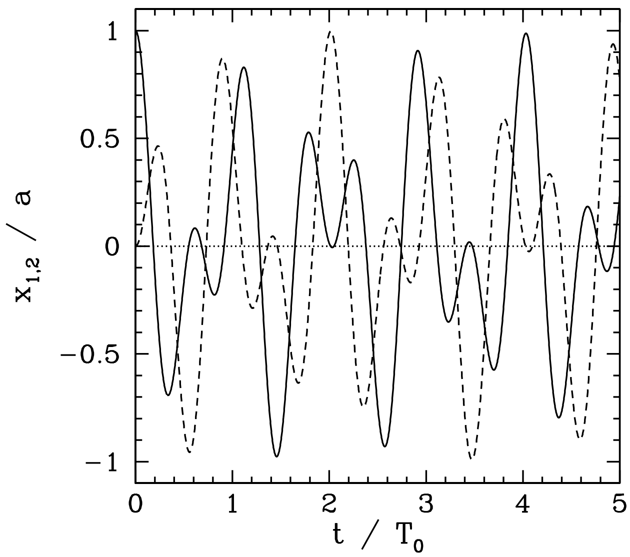


Figure 4.2: *Coupled oscillations in a two degree of freedom mass-spring system.*

Finally, let us define the so-called *normal coordinates*,

$$\eta_1(t) = [x_1(t) + x_2(t)]/2, \tag{4.28}$$

$$\eta_2(t) = [x_1(t) - x_2(t)]/2. \tag{4.29}$$

It follows from (4.20) and (4.21) that, in the presence of both normal modes,

$$\eta_1(t) = \hat{\eta}_1 \cos(\omega_0 t - \phi_1), \tag{4.30}$$

$$\eta_2(t) = \hat{\eta}_2 \cos(\sqrt{3}\,\omega_0 t - \phi_2). \tag{4.31}$$

Thus, in general, the two normal coordinates oscillate sinusoidally with *unique frequencies*, unlike the regular coordinates, $x_1(t)$ and $x_2(t)$—see Figure 4.2. This suggests that the equations of motion of the system should look particularly simple when expressed in terms of the normal coordinates. In fact, it is easily seen that the sum of Equations (4.3) and (4.4) reduces to

$$\ddot{\eta}_1 = -\omega_0^2\,\eta_1, \tag{4.32}$$

whereas the difference gives

$$\ddot{\eta}_2 = -3\,\omega_0^2\,\eta_2. \tag{4.33}$$

Thus, when expressed in terms of the normal coordinates, the equations of motion of the system reduce to two *uncoupled* simple harmonic oscillator equations. Of course, most general solution to Equation (4.32) is (4.30), whereas the most general solution to Equation (4.33) is (4.31). Hence, if we can guess the normal coordinates of a coupled oscillatory system then the determination of the normal modes of oscillation is considerably simplified.

4.2 Two Coupled LC Circuits

Consider the LC circuit pictured in Figure 4.3. Let $I_1(t)$, $I_2(t)$, and $I_3(t)$ be the currents flowing in the three legs of the circuit, which meet at junctions A and B. According to *Kichhoff's first circuital law*, the net current flowing into each junction is zero. It follows that $I_3 = -(I_1 + I_2)$. Hence, this is a *two degree of freedom* system whose instantaneous configuration is specified by the two independent variables $I_1(t)$ and $I_2(t)$. It follows that there are *two* independent normal modes of oscillation. Now, the potential differences across the left, middle, and right legs of the circuit are $Q_1/C + L\,\dot{I}_1$, Q_3/C', and $Q_2/C + L\,\dot{I}_2$, respectively, where $\dot{Q}_1 = I_1$, $\dot{Q}_2 = I_2$, and $Q_3 = -(Q_1 + Q_2)$. However, since the three legs are connected in *parallel*, the potential differences must all be equal, so that

$$Q_1/C + L\,\dot{I}_1 = Q_3/C' = -(Q_1 + Q_2)/C', \tag{4.34}$$

$$Q_2/C + L\,\dot{I}_2 = Q_3/C' = -(Q_1 + Q_2)/C'. \tag{4.35}$$

Differentiating with respect to t, and dividing by L, we obtain the coupled time evolution equations of the system:

$$\ddot{I}_1 + \omega_0^2\,(1+\alpha)\,I_1 + \omega_0^2\,\alpha\,I_2 = 0, \tag{4.36}$$

$$\ddot{I}_2 + \omega_0^2\,(1+\alpha)\,I_2 + \omega_0^2\,\alpha\,I_1 = 0, \tag{4.37}$$

where $\omega_0 = 1/\sqrt{L\,C}$ and $\alpha = C/C'$.

It is fairly easy to guess that the normal coordinates of the system are

$$\eta_1 = (I_1 + I_2)/2, \tag{4.38}$$

$$\eta_2 = (I_1 - I_2)/2. \tag{4.39}$$

Forming the sum and difference of Equations (4.36) and (4.37), we obtain the evolution equations for the two independent normal modes of oscillation:

$$\ddot{\eta}_1 + \omega_0^2\,(1 + 2\,\alpha)\,\eta_1 = 0, \tag{4.40}$$

$$\ddot{\eta}_2 + \omega_0^2\,\eta_2 = 0. \tag{4.41}$$

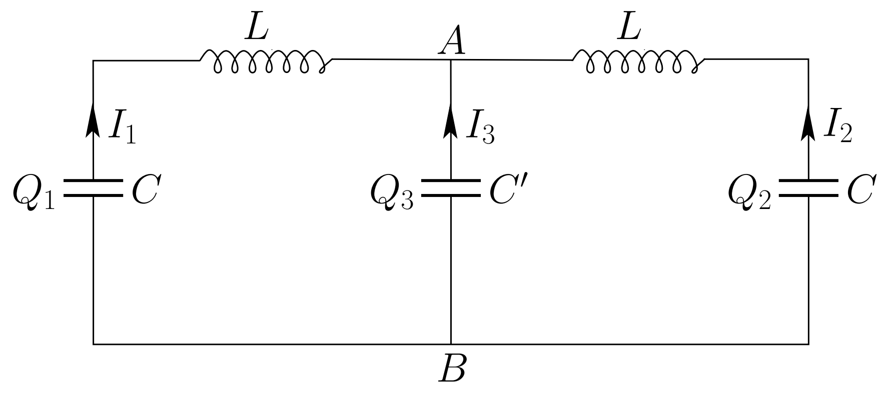


Figure 4.3: *Two degree of freedom* LC *circuit.*

These equations can readily be solved to give

$$\eta_1(t) = \hat{\eta}_1 \cos(\omega_1 t - \phi_1), \tag{4.42}$$

$$\eta_2(t) = \hat{\eta}_2 \cos(\omega_0 t - \phi_2), \tag{4.43}$$

where $\omega_1 = (1 + 2\alpha)^{1/2}\,\omega_0$. Here, $\hat{\eta}_1$, ϕ_1, $\hat{\eta}_2$, and ϕ_2 are constants determined by the initial conditions. It follows that

$$I_1(t) = \eta_1(t) + \eta_2(t) = \hat{\eta}_1 \cos(\omega_1 t - \phi_1) + \hat{\eta}_2 \cos(\omega_0 t - \phi_2), \tag{4.44}$$

$$I_2(t) = \eta_1(t) - \eta_2(t) = \hat{\eta}_1 \cos(\omega_1 t - \phi_1) - \hat{\eta}_2 \cos(\omega_0 t - \phi_2). \tag{4.45}$$

As an example, suppose that $\phi_1 = \phi_2 = 0$ and $\hat{\eta}_1 = \hat{\eta}_2 = I_0/2$. We obtain

$$I_1(t) = I_0 \cos(\omega_- t)\, \cos(\omega_+ t), \tag{4.46}$$

$$I_2(t) = I_0 \sin(\omega_- t)\, \sin(\omega_+ t), \tag{4.47}$$

where $\omega_\pm = (\omega_0 \pm \omega_1)/2$. This solution is illustrated in Figure 4.4. [Here, $T_0 = 2\pi/\omega_0$ and $\alpha = 0.2$. Thus, the two normal frequencies are ω_0 and $1.18\,\omega_0$.] Note the *beats* generated by the superposition of two normal modes with similar normal frequencies.

We can also solve the problem in a more systematic manner by specifically searching for a normal mode of the form

$$I_1(t) = \hat{I}_1 \cos(\omega t - \phi), \tag{4.48}$$

$$I_2(t) = \hat{I}_2 \cos(\omega t - \phi). \tag{4.49}$$

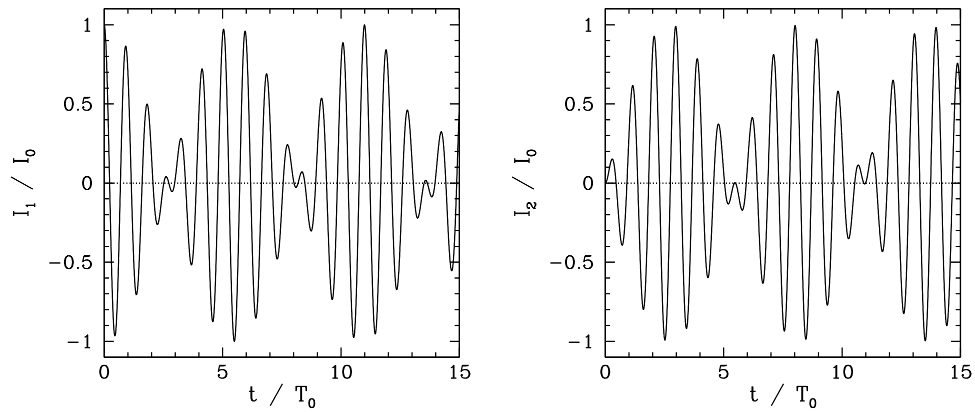


Figure 4.4: *Coupled oscillations in a two degree of freedom* LC *circuit.*

Substitution into the time evolution equations (4.36) and (4.37) yields the matrix equation

$$\begin{pmatrix} \hat{\omega}^2 - (1+\alpha) & -\alpha \\ -\alpha & \hat{\omega}^2 - (1+\alpha) \end{pmatrix} \begin{pmatrix} \hat{I}_1 \\ \hat{I}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{4.50}$$

where $\hat{\omega} = \omega/\omega_0$. The normal frequencies are determined by setting the determinant of the matrix to zero. This gives

$$\left[\hat{\omega}^2 - (1+\alpha)\right]^2 - \alpha^2 = 0, \tag{4.51}$$

or

$$\hat{\omega}^4 - 2\,(1+\alpha)\,\hat{\omega}^2 + 1 + 2\,\alpha = \left(\hat{\omega}^2 - 1\right)\left(\hat{\omega}^2 - [1+2\,\alpha]\right) = 0. \tag{4.52}$$

The roots of the above equation are $\hat{\omega} = 1$ and $\hat{\omega} = (1+2\,\alpha)^{1/2}$. (Again, we neglect the negative frequency roots, since they generate the same patterns of motion as the corresponding positive frequency roots.) Hence, the two normal frequencies are ω_0 and $(1+2\,\alpha)^{1/2}\,\omega_0$. The characteristic patterns of motion associated with the normal modes can be calculated from the first row of the matrix equation (4.50), which can be rearranged to give

$$\frac{\hat{I}_1}{\hat{I}_2} = \frac{\alpha}{\hat{\omega}^2 - (1+\alpha)}. \tag{4.53}$$

It follows that $\hat{I}_1 = -\hat{I}_2$ for the normal mode with $\hat{\omega} = 1$, and $\hat{I}_1 = \hat{I}_2$ for the normal mode with $\hat{\omega} = (1+2\,\alpha)^{1/2}$. We are thus led to Equations (4.44)–(4.45), where $\hat{\eta}_1$ and ϕ_1 are the amplitude and phase of the higher frequency normal mode, whereas $\hat{\eta}_2$ and ϕ_2 are the amplitude and phase of the lower frequency mode.

4.3 Three Spring Coupled Masses

Consider a generalized version of the mechanical system discussed in Section 4.1 that consists of *three* identical masses m which slide over a frictionless horizontal surface, and are connected by identical light horizontal springs of spring constant k. As before, the outermost masses are attached to immovable walls by springs of spring constant k. The instantaneous configuration of the system is specified by the horizontal displacements of the three masses from their equilibrium positions: *i.e.*, $x_1(t)$, $x_2(t)$, and $x_3(t)$. Clearly, this is a *three degree of freedom system*. We, therefore, expect it to possesses *three* independent normal modes of oscillation. Equations (4.1)–(4.2) generalize to

$$
\begin{aligned}
m\,\ddot{x}_1 &= -k\,x_1 + k\,(x_2 - x_1), &\text{(4.54)}\\
m\,\ddot{x}_2 &= -k\,(x_2 - x_1) + k\,(x_3 - x_2), &\text{(4.55)}\\
m\,\ddot{x}_3 &= -k\,(x_3 - x_2) + k\,(-x_3). &\text{(4.56)}
\end{aligned}
$$

These equations can be rewritten

$$
\begin{aligned}
\ddot{x}_1 &= -2\,\omega_0^2\,x_1 + \omega_0^2\,x_2, &\text{(4.57)}\\
\ddot{x}_2 &= \omega_0^2\,x_1 - 2\,\omega_0^2\,x_2 + \omega_0^2\,x_3, &\text{(4.58)}\\
\ddot{x}_3 &= \omega_0^2\,x_2 - 2\,\omega_0^2\,x_3, &\text{(4.59)}
\end{aligned}
$$

where $\omega_0 = \sqrt{k/m}$. Let us search for a normal mode solution of the form

$$
\begin{aligned}
x_1(t) &= \hat{x}_1\,\cos(\omega\,t - \phi), &\text{(4.60)}\\
x_2(t) &= \hat{x}_2\,\cos(\omega\,t - \phi), &\text{(4.61)}\\
x_3(t) &= \hat{x}_3\,\cos(\omega\,t - \phi). &\text{(4.62)}
\end{aligned}
$$

Equations (4.57)–(4.62) can be combined to give the 3×3 homogeneous matrix equation

$$
\begin{pmatrix} \hat{\omega}^{\,2} - 2 & 1 & 0 \\ 1 & \hat{\omega}^{\,2} - 2 & 1 \\ 0 & 1 & \hat{\omega}^{\,2} - 2 \end{pmatrix} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{4.63}
$$

where $\hat{\omega} = \omega/\omega_0$. The normal frequencies are determined by setting the determinant of the matrix to zero: *i.e.*,

$$(\hat{\omega}^2 - 2)\left[(\hat{\omega}^2 - 2)^2 - 1\right] - (\hat{\omega}^2 - 2) = 0, \tag{4.64}$$

or

$$(\hat{\omega}^2 - 2)\,\left[\hat{\omega}^2 - 2 - \sqrt{2}\right]\,\left[\hat{\omega}^2 - 2 + \sqrt{2}\right] = 0. \tag{4.65}$$

Thus, the normal frequencies are $\hat{\omega} = \sqrt{2}\,(1 - 1/\sqrt{2})^{1/2}$, $\sqrt{2}$, and $\sqrt{2}\,(1 + 1/\sqrt{2})^{1/2}$. According to the first and third rows of Equation (4.63),

$$\hat{x}_1 : \hat{x}_2 : \hat{x}_3 :: 1 : 2 - \hat{\omega}^2 : 1, \tag{4.66}$$

provided $\hat{\omega}^2 \neq 2$. According to the second row,

$$\hat{x}_1 : \hat{x}_2 : \hat{x}_3 :: -1 : 0 : 1 \tag{4.67}$$

when $\hat{\omega}^2 = 2$. Note that we can only determine the relative ratios of $\hat{x}_1$, $\hat{x}_2$, and $\hat{x}_3$, rather than the absolute values of these quantities. In other words, only the direction of the vector $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$ is well-defined. [This follows because the most general solution, (4.71), is undetermined to an arbitrary multiplicative constant. That is, if $\mathbf{x}(t) = (x_1(t), x_2(t), x_3(t))$ is a solution to the dynamical equations (4.57)–(4.59) then so is $a\,\mathbf{x}(t)$, where a is an arbitrary constant. This, in turn, follows because the dynamical equations are linear.] Let us arbitrarily set the magnitude of $\hat{\mathbf{x}}$ to unity. It follows that the normal mode associated with the normal frequency $\hat{\omega}_1 = \sqrt{2}\,(1 - 1/\sqrt{2})^{1/2}$ is

$$\hat{\mathbf{x}}_1 = \left(\frac{1}{2},\, \frac{1}{\sqrt{2}},\, \frac{1}{2}\right). \tag{4.68}$$

Likewise, the normal mode associated with the normal frequency $\hat{\omega}_2 = \sqrt{2}$ is

$$\hat{\mathbf{x}}_2 = \left(-\frac{1}{\sqrt{2}},\, 0,\, \frac{1}{\sqrt{2}}\right). \tag{4.69}$$

Finally, the normal mode associated with the normal frequency $\hat{\omega}_3 = \sqrt{2}(1 + 1/\sqrt{2})^{1/2}$ is

$$\hat{\mathbf{x}}_3 = \left(\frac{1}{2}, -\frac{1}{\sqrt{2}},\, \frac{1}{2}\right). \tag{4.70}$$

Let $\mathbf{x} = (x_1, x_2, x_2)$. It follows that the most general solution to the problem is

$$\mathbf{x}(t) = \eta_1(t)\,\hat{\mathbf{x}}_1 + \eta_2(t)\,\hat{\mathbf{x}}_2 + \eta_3(t)\,\hat{\mathbf{x}}_3, \tag{4.71}$$

where

$$\begin{aligned}\eta_1(t) &= \hat{\eta}_1 \cos(\hat{\omega}_1\, t - \phi_1), &(4.72)\\ \eta_2(t) &= \hat{\eta}_2 \cos(\hat{\omega}_2\, t - \phi_2), &(4.73)\\ \eta_3(t) &= \hat{\eta}_3 \cos(\hat{\omega}_3\, t - \phi_3). &(4.74)\end{aligned}$$

Here, $\hat{\eta}_{1,2,3}$ and $\phi_{1,2,3}$ are constants. Equation (4.71) yields

$$\begin{pmatrix} x_1\\ x_2\\ x_3\end{pmatrix} = \begin{pmatrix} 1/2 & 1/\sqrt{2} & 1/2\\ -1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 1/2 & -1/\sqrt{2} & 1/2\end{pmatrix}\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\end{pmatrix} \tag{4.75}$$

The above equation can easily be inverted by noting that the matrix is *unitary*: *i.e.*, its transpose is equal to its inverse. Thus, we obtain

$$\begin{pmatrix} \eta_1\\ \eta_2\\ \eta_3\end{pmatrix} = \begin{pmatrix} 1/2 & -1/\sqrt{2} & 1/2\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\\ 1/2 & 1/\sqrt{2} & 1/2\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix} \tag{4.76}$$

This equation determines the three normal coordinates, η_1, η_2, η_3, in terms of the three conventional coordinates, x_1, x_2, x_3. Note that, in general, the normal coordinates are undetermined to arbitrary multiplicative constants.

4.4 Exercises

1. A particle of mass m is attached to a rigid support by means of a spring of spring constant k. At equilibrium, the spring hangs vertically downward. An identical oscillator is added to this system, the spring of the former being attached to the mass of the latter. Calculate the normal frequencies for one-dimensional vertical oscillations, and describe the associated normal modes.

2. Consider a mass-spring system of the general form shown in Figure 4.1 in which the two masses are of mass m, the two outer springs have spring constant k, and the middle spring has spring constant k'. Find the normal frequencies and normal modes in terms of $\omega_0 = \sqrt{k/m}$ and $\alpha = k'/k$.

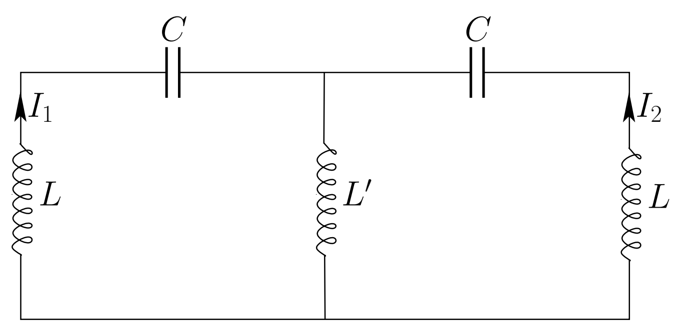
3. Consider a mass-spring system of the general form shown in Figure 4.1 in which the springs all have spring constant k, and the left and right masses are of mass m and m', respectively. Find the normal frequencies and normal modes in terms of $\omega_0 = \sqrt{k/m}$ and $\alpha = m'/m$.

4. Consider two simple pendula with the same length, l, but different bob masses, m_1 and m_2. Suppose that the pendula are connected by a spring of spring constant k. Let the spring be unextended when the two bobs are in their equilibrium positions. Demonstrate that the equations of motion of the system (for small amplitude oscillations) are

$$\begin{aligned} m_1\,\ddot{\theta}_1 &= -m_1\,\frac{g}{l}\,\theta_1 + k\,(\theta_2-\theta_1), \\ m_2\,\ddot{\theta}_2 &= -m_2\,\frac{g}{l}\,\theta_2 + k\,(\theta_2-\theta_1), \end{aligned}$$

where θ_1 and θ_2 are the angular displacements of the respective pendula from their equilibrium positions. Show that the normal coordinates are $\eta_1 = (m_1\,\theta_1 + m_2\,\theta_2)/(m_1+m_2)$ and $\eta_2 = \theta_1 - \theta_2$. Find the normal frequencies and normal modes. Find a superposition of the two modes such that at $t = 0$ the two pendula are stationary, with $\theta_1 = \theta_0$, and $\theta_2 = 0$.

5. Find the normal frequencies and normal modes of the coupled LC circuit shown below in terms of $\omega_0 = 1/\sqrt{L\,C}$ and $\alpha = L'/L$.



5 Transverse Standing Waves

5.1 Normal Modes of a Beaded String

Consider a mechanical system consisting of a taut string which is stretched between two immovable walls. Suppose that N identical beads of mass m are attached to the string in such a manner that they cannot slide along it. Let the beads be equally spaced a distance a apart, and let the distance between the first and the last beads and the neighboring walls also be a. See Figure 5.1. Consider *transverse* oscillations of the string: *i.e.*, oscillations in which the string moves from side to side (*i.e.*, in the y-direction). It is assumed that the inertia of the string is negligible with respect to that of the beads. It follows that the sections of the string between neighboring beads, and between the outermost beads and the walls, are *straight*. (Otherwise, there would be a net tension force acting on the sections, and they would consequently suffer an infinite acceleration.) In fact, we expect the instantaneous configuration of the string to be a set of continuous straight-line segments of varying inclinations, as shown in the figure. Finally, assuming that the transverse displacement of the string is *relatively small*, it is reasonable to suppose that each section of the string possesses the *same* tension, T.

It is convenient to introduce a Cartesian coordinate system such that x measure distance along the string from the left wall, and y measures the transverse displacement of the string from its equilibrium position. See Figure 5.1. Thus, when the string is in its equilibrium position it runs along the

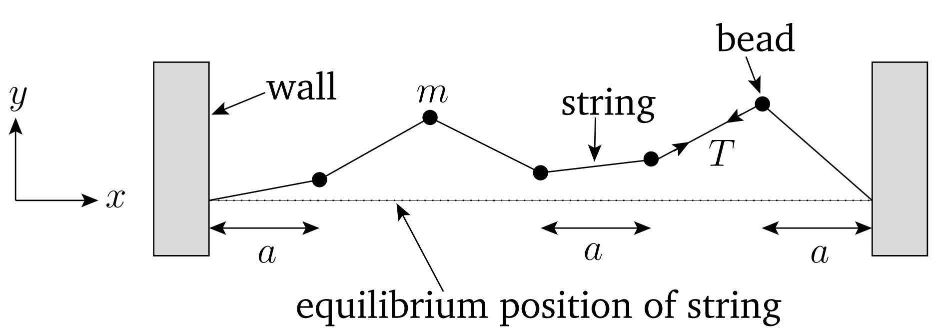


Figure 5.1: *A beaded string.*

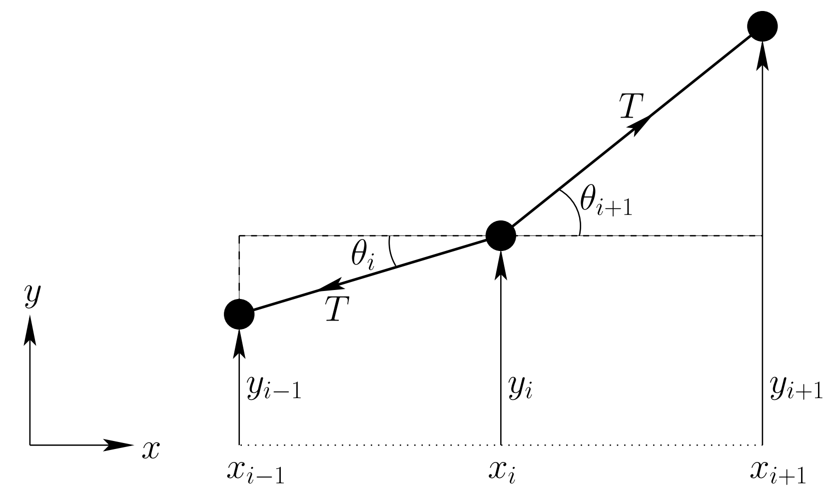


Figure 5.2: *A short section of a beaded string.*

x-axis. We can define

$$x_i = i\,a, \tag{5.1}$$

where $i = 1, 2, \cdots, N$. Here, x_1 is the x-coordinate of the closest bead to the left wall, x_2 the x-coordinate of the second closest bead, *etc.* The x-coordinates of the beads are assumed to remain *constant* during the transverse oscillations. We can also define $x_0 = 0$ and $x_{N+1} = (N+1)\,a$ as the x-coordinates of the left and right ends of the string. Let the transverse displacement of the ith bead be $y_i(t)$, for $i = 1, N$. Since each displacement can vary independently, we are clearly dealing with an N degree of freedom system. We would, therefore, expect such a system to possess N unique normal modes of oscillation.

Consider the section of the string lying between the $i-1$th and $i+1$th beads, as shown in Figure 5.2. Here, $x_{i-1} = x_i - a$, x_i, and $x_{i+1} = x_i + a$ are the distances of the $i-1$th, ith, and $i+1$th beads, respectively, from the left wall, whereas y_{i+1}, y_i, and y_{i+1} are the corresponding transverse displacements of these beads. The two sections of the string which are attached to the ith bead subtend angles θ_i and θ_{i+1} with the x-axis, as illustrated in the figure. Simple trigonometry reveals that

$$\tan\theta_i = \frac{y_i - y_{i-1}}{x_i - x_{i-1}} = \frac{y_i - y_{i-1}}{a}, \tag{5.2}$$

and

$$\tan\theta_{i+1} = \frac{y_{i+1} - y_i}{a}. \tag{5.3}$$

However, if the transverse displacement of the string is relatively small—*i.e.*, if $|y_i| \ll a$ for all i—which we shall assume to be the case, then θ_i and θ_{i+1} are both *small* angles. Now, it is well known that $\tan\theta \simeq \theta$ when $|\theta| \ll 1$. It follows that

$$\theta_i \simeq \frac{y_i - y_{i-1}}{a}, \tag{5.4}$$

$$\theta_{i+1} \simeq \frac{y_{i+1} - y_i}{a}. \tag{5.5}$$

Let us find the transverse equation of motion of the ith bead. This bead is subject to two forces: *i.e.*, the tensions in the sections of the string to the left and to the right of it. These tensions are of magnitude T, and are directed parallel to the associated string sections, as shown in Figure 5.2. Thus, the transverse (*i.e.*, y-directed) components of these two tensions are $-T\,\sin\theta_i$ and $T\,\sin\theta_{i+1}$, respectively. Hence, the transverse equation of motion of the ith bead becomes

$$m\,\ddot{y}_i = -T\,\sin\theta_i + T\,\sin\theta_{i+1}. \tag{5.6}$$

However, since θ_i and θ_{i+1} are both small angles, we can employ the small angle approximation $\sin\theta \simeq \theta$. It follows that

$$\ddot{y}_i \simeq \frac{T}{m}\left(\theta_{i+1} - \theta_i\right). \tag{5.7}$$

Finally, making use of Equations (5.4) and (5.5), we obtain

$$\ddot{y}_i = \omega_0^{\,2}\left(y_{i-1} - 2\,y_i + y_{i+1}\right), \tag{5.8}$$

where $\omega_0 = \sqrt{T/m\,a}$. Since there is nothing special about the ith bead, we deduce that the above equation of motion applies to all N beads: *i.e.*, it is valid for $i = 1, N$. Of course, the first ($i = 1$) and last ($i = N$) beads are special cases, since there is no bead corresponding to $i = 0$ or $i = N+1$. In fact, $i = 0$ and $i = N+1$ correspond to the left and right ends of the string, respectively. However, Equation (5.8) still applies to the first and last beads as long as we set $y_0 = 0$ and $y_{N+1} = 0$. What we are effectively demanding is that the two ends of the string, which are attached to the left and right walls, must both have *zero transverse displacement*.

Let us search for a normal mode solution to Equation (5.8) which takes the form

$$y_i(t) = A\,\sin(k\,x_i)\,\cos(\omega\,t - \phi), \tag{5.9}$$

where $A > 0$, $k > 0$, $\omega > 0$, and ϕ are constants. This particular type of solution is such that all of the beads execute transverse simple harmonic oscillations *in phase* with one another. See Figure 5.4. Moreover, the oscillations have an amplitude $A\,\sin(k\,x_i)$ which *varies sinusoidally* along the length of the string (*i.e.*, in the x-direction). The pattern of oscillations is thus *periodic in space*. The spatial repetition period, which is usually known as the *wavelength*, is $\lambda = 2\pi/k$. [This follows from (5.9) because $\sin\theta$ is a periodic function with period 2π: *i.e.*, $\sin(\theta + 2\pi) \equiv \sin\theta$.] The constant k, which determines the wavelength, is usually referred to as the *wavenumber*. Thus, a small wavenumber corresponds to a long wavelength, and *vice versa*. The type of solution specified in (5.9) is generally known as a *standing wave*. It is a *wave* because it is periodic in both space and time. (An oscillation is periodic in time only.) It is a *standing* wave, rather than a traveling wave, because the points of maximum and minimum amplitude oscillation are stationary (in x). See Figure 5.4.

Substituting (5.9) into (5.8), we obtain

$$\begin{aligned} -\,\omega^2 A\,\sin(k\,x_i)\,\cos(\omega\,t-\phi) &= \omega_0^2\,A\,[\sin(k\,x_{i-1}) - 2\,\sin(k\,x_i) \\ &\quad + \sin(k\,x_{i+1})]\,\cos(\omega\,t-\phi), \end{aligned} \tag{5.10}$$

which yields

$$-\,\omega^2\,\sin(k\,x_i) = \omega_0^2\,(\sin[k\,(x_i - a)] - 2\,\sin(k\,x_i) + \sin[k\,(x_i + a)])\,. \tag{5.11}$$

However, since $\sin(a+b) \equiv \sin a\,\cos b + \cos a\,\sin b$, we get

$$-\,\omega^2\,\sin(k\,x_i) = \omega_0^2\,[\cos(k\,a) - 2 + \cos(k\,a)]\,\sin(k\,x_i), \tag{5.12}$$

or

$$\omega^2 = 4\,\omega_0^2\,\sin^2(k\,a/2), \tag{5.13}$$

where use has been made of the trigonometric identity $1-\cos\theta \equiv 2\,\sin^2(\theta/2)$. Note that an expression, such as (5.13), which determines the angular frequency, ω, of a wave in terms of its wavenumber, k, is generally known as a *dispersion relation*.

Now, the solution (5.9) is only physical provided $y_0 = y_{N+1} = 0$: *i.e.*, provided that the two ends of the string remain stationary. The first constraint is automatically satisfied, since $x_0 = 0$ [see (5.1)]. The second constraint implies that

$$\sin(k\,x_{N+1}) = \sin[(N+1)\,k\,a] = 0. \tag{5.14}$$

This condition can only be satisfied if

$$k = \frac{n}{N+1}\,\frac{\pi}{a}, \tag{5.15}$$

where the integer n is known as the *mode number*. Clearly, a small mode number translates to a small wavenumber, and, hence, to a long wavelength, and *vice versa*. We conclude that the possible wavenumbers, k, of the normal modes of the system are *quantized* such that they are integer multiples of $\pi/[(N+1)\,a]$. Thus, the nth normal mode is associated with the characteristic pattern of bead displacements

$$y_{n,i}(t) = A_n \sin\left(\frac{n\,i}{N+1}\,\pi\right)\cos(\omega_n\,t - \phi_n), \tag{5.16}$$

where

$$\omega_n = 2\,\omega_0 \sin\left(\frac{n}{N+1}\,\frac{\pi}{2}\right). \tag{5.17}$$

Here, the integer $i = 1, N$ indexes the beads, whereas the mode number n indexes the normal modes. Furthermore, A_n and ϕ_n are arbitrary constants determined by the initial conditions. Of course, A_n is the *peak amplitude* of the nth normal mode, whereas ϕ_n is the associated *phase angle*.

So, how many unique normal modes does the system possess? At first sight, it might seem that there are an infinite number of normal modes, corresponding to the infinite number of possible values that the integer n can take. However, this is not the case. For instance, if $n = 0$ or $n = N+1$ then all of the $y_{n,i}$ are zero. Clearly, these cases are not real normal modes. Moreover, it is easily demonstrated that

$$\omega_{-n} = -\omega_n, \tag{5.18}$$

$$y_{-n,i}(t) = y_{n,i}(t), \tag{5.19}$$

provided $A_{-n} = -A_n$ and $\phi_{-n} = -\phi_n$, as well as

$$\omega_{N+1+n} = \omega_{N+1-n}, \tag{5.20}$$

$$y_{N+1+n,i}(t) = y_{N+1-n,i}(t), \tag{5.21}$$

provided $A_{N+1+n} = -A_{N+1-n}$ and $\phi_{N+1+n} = \phi_{N+1-n}$. We, thus, conclude that only those normal modes which have n in the range 1 to N correspond to unique modes. Modes with n values lying outside this range are either null modes, or modes that are identical to other modes with n values lying within the prescribed range. It follows that there are N unique normal modes of the form (5.16). Hence, given that we are dealing with an N degree of freedom system, which we would expect to only possess N unique normal modes, we can be sure that we have found *all* of the normal modes of the system.

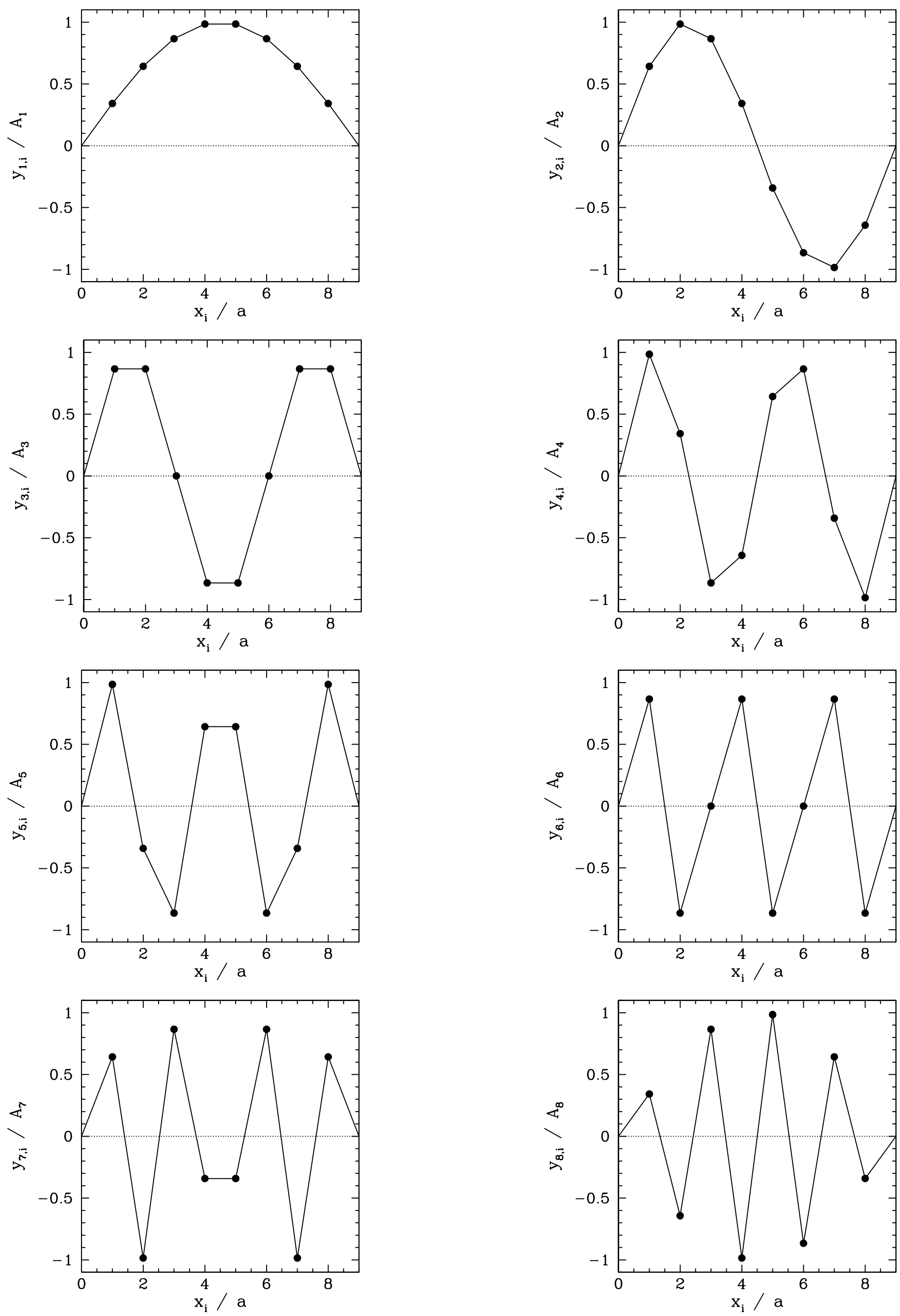


Figure 5.3: *Normal modes of a beaded string with eight equally spaced beads.*

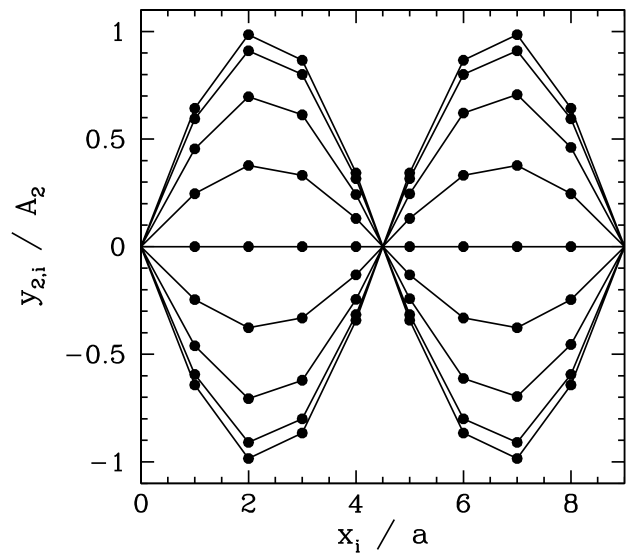


Figure 5.4: *Time evolution of the $n = 2$ normal mode of a beaded string with eight equally spaced beads.*

Figure 5.3 illustrates the spatial variation of the normal modes of a beaded string possessing eight beads: *i.e.*, an $N = 8$ system. The modes are all shown at the instances in time at which they attain their maximum amplitudes: *i.e.*, at $\omega_n t - \phi_n = 0$. It can be seen that the low-n—*i.e.*, long wavelength—modes cause the string to oscillate in a fairly smoothly varying (in x) sine wave pattern. On the other hand, the high-n—*i.e.*, short wavelength—modes cause the string to oscillate in a rapidly varying zig-zag pattern which bears little resemblance to a sine wave. The crucial distinction between the two different types of mode is that the wavelength of the oscillation (in the x-direction) is much larger than the bead spacing in the former case, whilst it is similar to the bead spacing in the latter. For instance, $\lambda = 18\,a$ for the $n = 1$ mode, $\lambda = 9\,a$ for the $n = 2$ mode, but $\lambda = 2.25\,a$ for the $n = 8$ mode.

Figure 5.5 displays the temporal variation of the $n = 2$ normal mode of an $N = 8$ beaded string. The mode is shown at $\omega_2 t - \phi_2 = 0$, $\pi/8$, $\pi/4$, $3\pi/8$, $\pi/2$, $5\pi/8$, $3\pi/2$, $7\pi/8$ and π. It can be seen that the beads oscillate *in phase* with one another: *i.e.*, they all attain their maximal transverse displacements, and pass through zero displacement, *simultaneously*. Note that the mid-way point of the string always remains stationary. Such a point

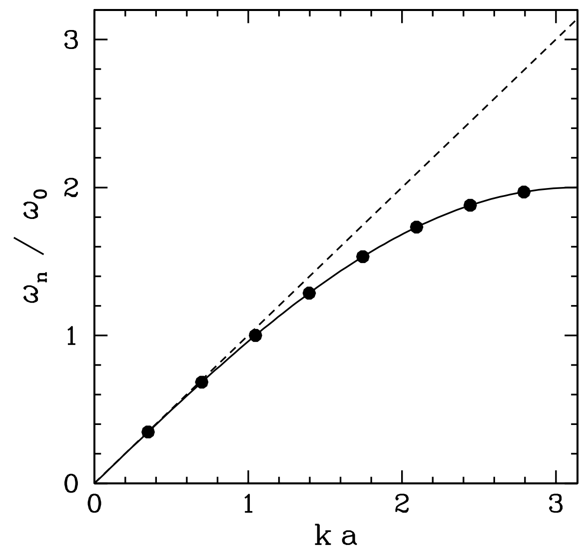


Figure 5.5: *Normal frequencies of a beaded string with eight equally spaced beads.*

is known as a *node*. The $n = 1$ normal mode has two nodes (counting the stationary points at each end of the string as nodes), the $n = 2$ mode has three nodes, the $n = 2$ mode four nodes, *etc.* In fact, the existence of nodes is one of the distinguishing features of a standing wave.

Figure 5.5 shows the normal frequencies of an $N = 8$ beaded string plotted as a function of the normalized wavenumber. Recall that, for an $N = 8$ system, the relationship between the normalized wavenumber, $k\,a$, and the mode number, n, is $k\,a = (n/9)\,\pi$. It can be seen that the angular frequency increases as the wavenumber increases, which implies that shorter wavelength modes have higher oscillation frequencies. Note that the dependence of the angular frequency on the normalized wavenumber, $k\,a$, is approximately *linear* when $k\,a \ll 1$. Indeed, it can be seen from Equation (5.17) that if $k\,a \ll 1$ then the small angle approximation $\sin\theta \simeq \theta$ yields a linear dispersion relation of the form

$$\omega_n \simeq (k\,a)\,\omega_0 = \left(\frac{n}{N+1}\right)\pi\,\omega_0. \tag{5.22}$$

We, thus, conclude that those normal modes of a uniformly beaded string whose wavelengths greatly exceed the bead spacing (*i.e.*, modes with $k\,a \ll 1$) have approximately *linear dispersion relations* in which their angular frequencies are directly proportional to their mode numbers. However, it is clear from the figure that this linear relationship breaks down as $k\,a \to 1$, and the mode wavelength consequently becomes similar to the bead spacing.

5.2 Normal Modes of a Uniform String

Consider a uniformly beaded string in the limit in which the number of beads, N, becomes increasingly large, whilst the spacing, a, and the individual mass, m, of the beads becomes increasingly small. Let the limit be taken in such a manner that the length, $l = (N+1)\,a$, and the average mass per unit length, $\rho = m/a$, of the string both remain constant. Clearly, as N increases, and becomes very large, such a string will more and more closely approximate a *uniform string* of length l and mass per unit length ρ. What can we guess about the normal modes of a uniform string from the analysis contained in the previous section? Well, first of all, we would guess that a uniform string is an *infinite* degree of freedom system, with an infinite number of unique normal modes of oscillation. This follows because a uniform string is the $N \to \infty$ limit of a beaded string, and a beaded string possesses N unique normal modes. Next, we would guess that the normal modes of a uniform string exhibit *smooth sinusoidal spatial variation* in the x-direction, and that the angular frequency of the modes is *directly proportional* to their wavenumber. These last two conclusions follow because all of the normal modes of a beaded string are characterized by $k\,a \ll 1$ in the limit in which the spacing between the beads becomes zero. Let us now investigate whether these guesses are correct.

Consider the transverse oscillations of a uniform string of length l and mass per unit length ρ which is stretched between two immovable walls. It is again convenient to define a Cartesian coordinate system in which x measures distance along the string from the left wall, and y measures the transverse displacement of the string. Thus, the instantaneous state of the system at time t is determined by the function $y(x,t)$ for $0 \leq x \leq l$. Of course, this function consists of an infinite number of different y values, corresponding to the infinite number of different x values in the range 0 to l. Moreover, all of these y values are free to vary *independently* of one another. It follows that we are indeed dealing with a dynamical system possessing an *infinite* number of degrees of freedom.

Let us try to reuse some of the analysis of the previous section. We can reinterpret $y_i(t)$ as $y(x,t)$, $y_{i-1}(t)$ as $y(x-\delta x,t)$, and $y_{i+1}(t)$ as $y(x+\delta x,t)$, assuming that $x_i = x$ and $a = \delta x$. Moreover, $\ddot{y}_i(t)$ becomes $\partial^2 y(x,t)/\partial t^2$: *i.e.*, a second derivative of $y(x,t)$ with respect to t at *constant* x. Finally, $\omega_0^2 = T/(m\,a)$, where T is the tension in the string, can be rewritten as $T/[\rho\,(\delta x)^2]$, since $\rho = m/\delta x$. Incidentally, we are again assuming that the transverse displacement of the string remains sufficiently small that the tension is approximately constant in x. Thus, the equation of motion of the beaded string, (5.8), transforms into

$$\frac{\partial^2 y(x,t)}{\partial t^2} = \frac{T}{\rho}\left[\frac{y(x-\delta x,t) - 2\,y(x,t) + y(x+\delta x,t)}{(\delta x)^2}\right]. \tag{5.23}$$

However, Taylor expanding $y(x+\delta x,t)$ in x at constant t, we obtain

$$y(x+\delta x,t) = y(x,t) + \frac{\partial y(x,t)}{\partial x}\,\delta x + \frac{1}{2}\,\frac{\partial^2 y(x,t)}{\partial x^2}\,(\delta x)^2 + \mathcal{O}(\delta x)^3. \tag{5.24}$$

Likewise,

$$y(x-\delta x,t) = y(x,t) - \frac{\partial y(x,t)}{\partial x}\,\delta x + \frac{1}{2}\,\frac{\partial^2 y(x,t)}{\partial x^2}\,(\delta x)^2 + \mathcal{O}(\delta x)^3. \tag{5.25}$$

It follows that

$$\left[\frac{y(x-\delta x,t) - 2\,y(x,t) + y(x+\delta x,t)}{(\delta x)^2}\right] = \frac{\partial^2 y(x,t)}{\partial x^2} + \mathcal{O}(\delta x). \tag{5.26}$$

Thus, in the limit that $\delta x \to 0$, Equation (5.23) reduces to

$$\frac{\partial^2 y}{\partial t^2} = v^2\,\frac{\partial^2 y}{\partial x^2}, \tag{5.27}$$

where

$$v = \sqrt{\frac{T}{\rho}} \tag{5.28}$$

is a quantity having the dimensions of *velocity*. Equation (5.27), which is the transverse equation of motion of the string, takes the form of a very famous partial differential equation known as the *wave equation*. The quantity v turns out to the the propagation velocity of transverse waves along the string. See Section 7.1.

By analogy with Equation (5.9), let us search for a solution of the wave equation of the form

$$y(x,t) = A\,\sin(k\,x)\,\cos(\omega\,t - \phi), \tag{5.29}$$

where $A > 0$, $k > 0$, $\omega > 0$, and ϕ are constants. We would interpret such a solution as a *standing wave* of wavenumber k, wavelength $\lambda = 2\pi/k$, angular frequency ω, peak amplitude A, and phase angle ϕ. Substitution of the above expression into Equation (5.27) yields the *dispersion relation* [*cf.*, (5.13)]

$$\omega^2 = k^2 v^2. \tag{5.30}$$

Now, the standing wave solution (5.29) is subject to the physical constraint that the two ends of the string, which are attached to immovable walls, remain *stationary*. This leads directly to the *boundary conditions*

$$y(0,t) = 0, \tag{5.31}$$

$$y(l,t) = 0. \tag{5.32}$$

It can be seen that the solution (5.29) automatically satisfies the first boundary condition. However, the second boundary condition is only satisfied when $\sin(k\,l) = 0$, which immediately implies that

$$k = n\,\frac{\pi}{l}, \tag{5.33}$$

where the mode number, n, is an integer. We, thus, conclude that the possible normal modes of a taut string, of length l and fixed ends, have wavenumbers which are *quantized* such that they are integer multiples of π/l. Moreover, it is clear that this quantization is a direct consequence of the imposition of the physical boundary conditions at the two ends of the string.

It follows, from the above analysis, that the nth normal mode of the string is associated with the pattern of motion

$$y_n(x,t) = A_n \sin\left(n\,\pi\,\frac{x}{l}\right)\cos(\omega_n\,t - \phi_n), \tag{5.34}$$

where

$$\omega_n = n\,\frac{\pi\,v}{l}. \tag{5.35}$$

Here, A_n and ϕ_n are constants which are determined by the initial conditions. See Section 5.3. So, how many unique normal modes are there? Well, the choice $n = 0$ yields $y_0(x,t) = 0$ for all x and t, so this is not a real normal mode. Moreover,

$$\omega_{-n} = -\omega_n, \tag{5.36}$$

$$y_{-n}(x,t) = y_n(x,t), \tag{5.37}$$

provided that $A_{-n} = -A_n$ and $\phi_{-n} = -\phi_n$. We, thus, conclude that modes with negative mode numbers give rise to the same patterns of motion as modes with corresponding positive mode numbers. However, the modes with positive mode numbers each correspond to unique patterns of motion which oscillate at unique frequencies. It follows that the string possesses an *infinite* number of normal modes, corresponding to the mode numbers $n = 1, 2, 3$, *etc.* Recall that we are dealing with an infinite degree of freedom system, which we would expect to possess an infinite number of unique normal modes. The fact that we have actually found an infinite number of such modes suggests that we have found *all* of the normal modes of the system.

Figure 5.6 illustrates the spatial variation of the first eight normal modes of a uniform string with fixed ends. The modes are all shown at the instances in time at which they attain their maximum amplitudes: *i.e.*, at $\omega_n t - \phi_n = 0$. It can be seen that the modes are all smoothly varying sine waves. The low mode number (*i.e.*, long wavelength) modes are actually quite similar in form to the corresponding normal modes of a uniformly beaded string. See Figure 5.3. However, the high mode number modes are substantially different. We conclude that the normal modes of a beaded string are similar to those of a uniform string, with the same length and mass per unit length, provided that the wavelength of the mode is much larger than the spacing between the beads.

Figure 5.7 illustrates the temporal variation of the $n = 4$ normal mode of a uniform string. The mode is shown at $\omega_4 t - \phi_4 = 0$, $\pi/8$, $\pi/4$, $3\pi/8$, $\pi/2$, $5\pi/8$, $3\pi/2$, $7\pi/8$ and π. It can be seen that all points on the string attain their maximal transverse displacements, and pass through zero displacement, *simultaneously*. Note that the $n = 4$ mode possesses *five* nodes, at which the string remains stationary. Two of these are located at the ends of the string, and three in the middle. In fact, it is clear from Equation (5.34) that the nodes correspond to points at which $\sin[n\,(x/l)\,\pi] = 0$. Hence, the nodes are located at

$$x_{n,j} = \left(\frac{j}{n}\right) l, \tag{5.38}$$

where j is an integer lying in the range 0 to n. Here, n indexes the normal mode, and j the node. Thus, the $j = 0$ node lies at the left end of the string, the $j = 1$ node is the next node to the right, *etc.* It is apparent, from the above formula, that the nth normal mode has $n + 1$ nodes which are *uniformly spaced* a distance l/n apart.

Finally, Figure 5.8 shows the first eight normal frequencies of a uniform string with fixed ends, plotted as a function of the mode number. It can

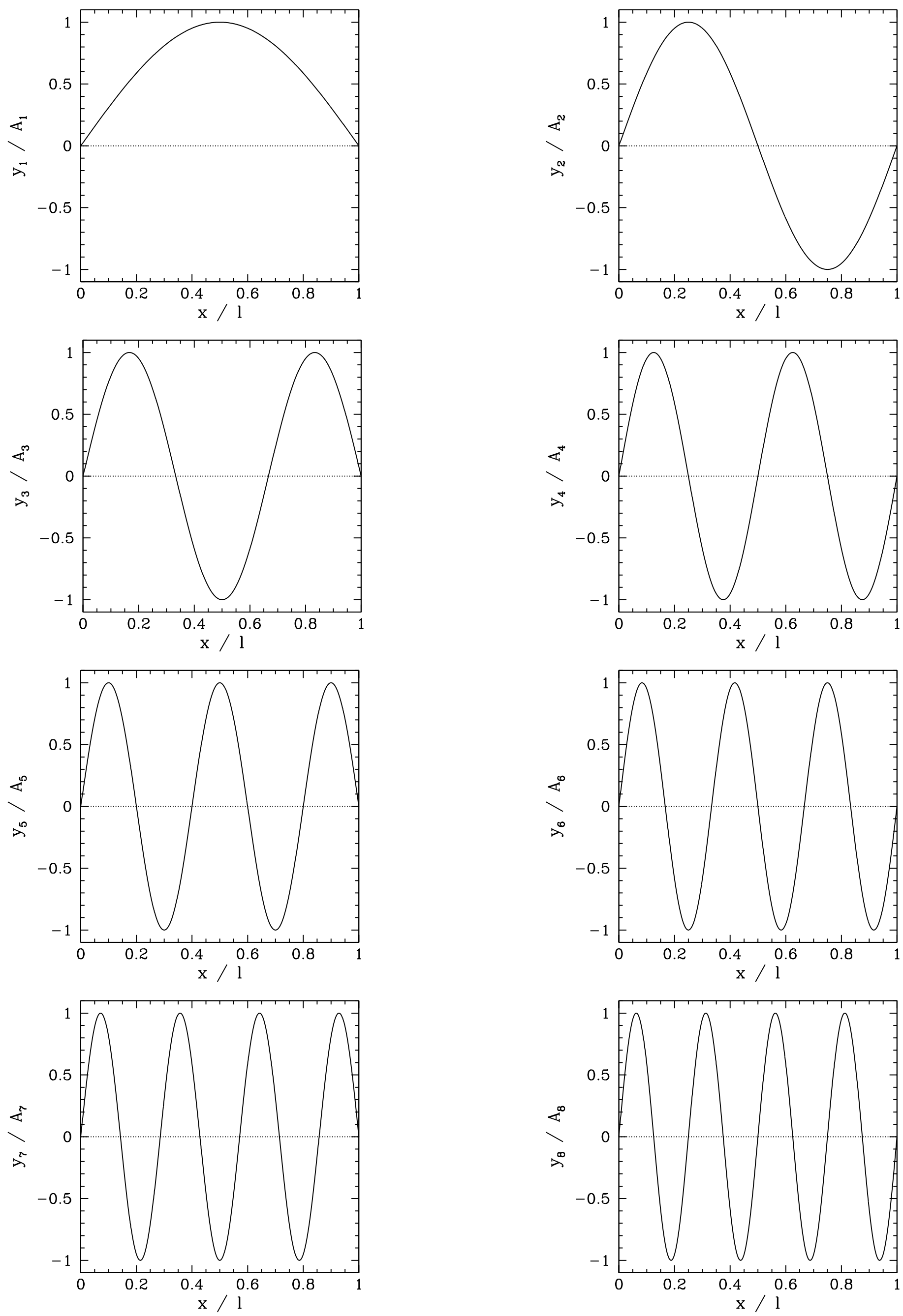


Figure 5.6: *First eight normal modes of a uniform string.*

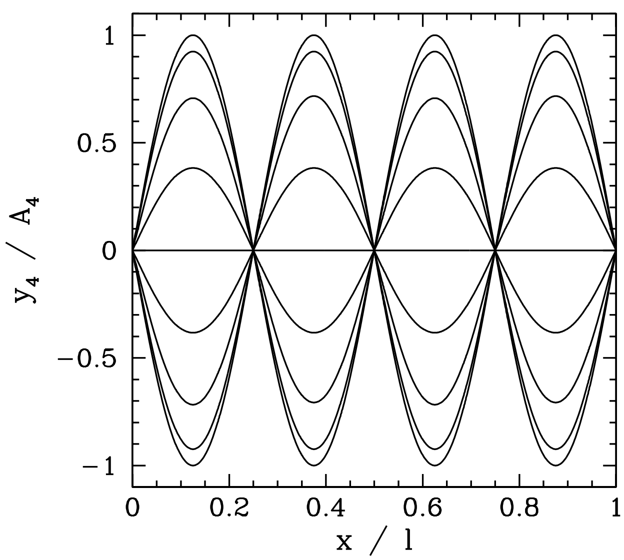


Figure 5.7: *Time evolution of the* $n = 4$ *normal mode of a uniform string.*

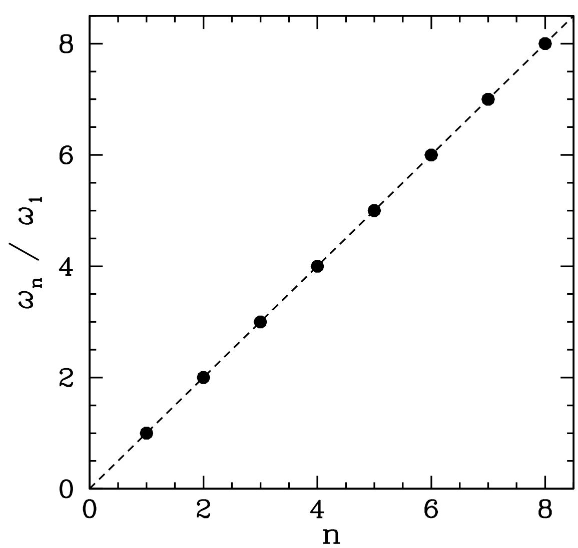


Figure 5.8: *Normal frequencies of the first eight normal modes of a uniform string.*

be seen that the angular frequency of oscillation increases *linearly* with the mode number. Recall that the low mode number (*i.e.*, long wavelength) normal modes of a beaded string also exhibit a linear relationship between normal frequency and mode number of the form [see Equation (5.22)]

$$\omega_n = \frac{n\,\pi}{N+1}\,\omega_0 = \frac{n\,\pi}{N+1}\left(\frac{T}{m\,a}\right)^{1/2}. \tag{5.39}$$

However, $m = \rho\,a$ and $l = (N+1)\,a$, so we obtain

$$\omega_n = \frac{n\,\pi}{l}\left(\frac{T}{\rho}\right)^{1/2} = n\,\frac{\pi\,v}{l}, \tag{5.40}$$

which is identical to Equation (5.35). We, thus, conclude that the normal frequencies of a uniformly beaded string are similar to those of a uniform string, with the same length and mass per unit length, as long as the wavelength of the associated normal mode is much larger than the spacing between the beads.

5.3 General Time Evolution of a Uniform String

In the previous section, we found the normal modes of a uniform string of length l, both ends of which are attached to immovable walls. These modes are spatially periodic solutions of the wave equation (5.27) which oscillate at unique frequencies and satisfy the boundary conditions (5.31) and (5.32). Since the wave equation is obviously *linear* [*i.e.*, if $y(x,t)$ is a solution then so is $a\,y(x,t)$, where a is an arbitrary constant], it follows that its most general solution is a *linear combination* of all of the normal modes. In other words,

$$y(x,t) = \sum_{n'=1,\infty} y_{n'}(x,t) = \sum_{n'=1,\infty} A_{n'}\,\sin\left(n'\,\pi\,\frac{x}{l}\right)\,\cos\left(n'\,\pi\,\frac{v\,t}{l} - \phi_{n'}\right), \tag{5.41}$$

where use has been made of (5.34) and (5.35). Note that this expression is obviously a solution of (5.27), and also automatically satisfies the boundary conditions (5.31) and (5.32). As we have already mentioned, the constants A_n and ϕ_n are determined by the *initial conditions*. Let us see how this comes about in more detail.

Suppose that the initial displacement of the string at $t = 0$ is

$$y_0(x) \equiv y(x,0). \tag{5.42}$$

Likewise, let the initial velocity of the string be

$$v_0(x) \equiv \frac{\partial y(x,0)}{\partial t}. \tag{5.43}$$

Obviously, for consistency with the boundary conditions, we must have $y_0(0) = y_0(l) = v_0(0) = v_0(l) = 0$. It follows from Equation (5.41) that

$$\begin{aligned} y_0(x) &= \sum_{n'=1,\infty} A_{n'} \cos\phi_{n'} \sin\left(n'\pi\frac{x}{l}\right), & (5.44)\\ v_0(x) &= \frac{v}{l}\sum_{n'=1,\infty} n'\pi A_{n'} \sin\phi_{n'} \sin\left(n'\pi\frac{x}{l}\right). & (5.45)\end{aligned}$$

Now, it is readily demonstrated that

$$\begin{aligned} &\frac{2}{l}\int_0^l \sin\left(n\pi\frac{x}{l}\right) \sin\left(n'\pi\frac{x}{l}\right) dx \\ &= \frac{2}{\pi}\int_0^\pi \sin\left(n\theta\right) \sin\left(n'\theta\right) d\theta \\ &= \frac{1}{\pi}\int_0^\pi \cos\left[(n-n')\,\theta\right] d\theta - \frac{1}{\pi}\int_0^l \cos\left[(n+n')\,\theta\right] d\theta, \\ &= \frac{1}{\pi}\left[\frac{\sin\left[(n-n')\,\theta\right]}{n-n'} - \frac{\sin\left[(n+n')\,\theta\right]}{n+n'}\right]_0^\pi \\ &= \frac{\sin\left[(n-n')\,\pi\right]}{(n-n')\,\pi} - \frac{\sin\left[(n+n')\,\pi\right]}{(n+n')\,\pi}, \end{aligned} \tag{5.46}$$

where n and n' are (possibly different) positive integers, $\theta = \pi x/l$, and use has been made of the trigonometric identity $2\sin a \sin b \equiv \cos(a-b) - \cos(a+b)$. Furthermore, it is easily seen that if k is a *non-zero* integer then

$$\frac{\sin(k\pi)}{k\pi} = 0. \tag{5.47}$$

On the other hand, $k = 0$ is a special case, since both the numerator and the denominator in the above expression become zero simultaneously. However, application of *l'Hopital's rule* yields

$$\lim_{x\to 0}\frac{\sin x}{x} = \lim_{x\to 0}\frac{d(\sin x)/dx}{dx/dx} = \lim_{x\to 0}\frac{\cos x}{1} = 1. \tag{5.48}$$

It follows that

$$\frac{\sin(k\pi)}{k\pi} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}, \tag{5.49}$$

where k is *any* integer. This result can be combined with Equation (5.46), recalling that n and n' are both *positive* integers, to give

$$\frac{2}{l}\int_0^l \sin\left(n\,\pi\,\frac{x}{l}\right)\,\sin\left(n'\,\pi\,\frac{x}{l}\right)\,dx = \delta_{n,n'}. \tag{5.50}$$

Here, the quantity

$$\delta_{n,n'} = \left\{\begin{array}{ll} 1 & n = n' \\ 0 & n \neq n' \end{array}\right., \tag{5.51}$$

where n and n' are integers, is known as the *Kronecker delta function*.

Let us multiply Equation (5.44) by $(2/l)\,\sin(n\,\pi\,x/l)$ and integrate over x from 0 to l. We obtain

$$\begin{aligned}
&\frac{2}{l}\int_0^l y_0(x)\,\sin\left(n\,\pi\,\frac{x}{l}\right)\,dx \\
&= \sum_{n'=1,\infty} A_{n'}\,\cos\phi_{n'}\,\frac{2}{l}\int_0^\pi \sin\left(n'\,\pi\,\frac{x}{l}\right)\,\sin\left(n\,\pi\,\frac{x}{l}\right)\,dx \\
&= \sum_{n'=1,\infty} A_{n'}\,\cos\phi_{n'}\,\delta_{n,n'} = A_n\cos\phi_n,
\end{aligned} \tag{5.52}$$

where use has been made of Equations (5.50) and (5.51). Similarly, Equation (5.45) yields

$$\frac{2}{v}\int_0^l v_0(x)\,\sin\left(n\,\pi\,\frac{x}{l}\right)\,dx = n\,\pi\,A_n\,\sin\phi_n. \tag{5.53}$$

Thus, defining the integrals

$$C_n = \frac{2}{l}\int_0^l y_0(x)\,\sin\left(n\,\pi\,\frac{x}{l}\right)\,dx, \tag{5.54}$$

$$S_n = \frac{2}{n\,\pi\,v}\int_0^l v_0(x)\,\sin\left(n\,\pi\,\frac{x}{l}\right)\,dx, \tag{5.55}$$

for $n = 1, \infty$, we obtain

$$C_n = A_n\,\cos\phi_n, \tag{5.56}$$

$$S_n = A_n\,\sin\phi_n, \tag{5.57}$$

and, hence,

$$A_n = (C_n^{\,2} + S_n^{\,2})^{1/2}, \tag{5.58}$$

$$\phi_n = \tan^{-1}(S_n/C_n). \tag{5.59}$$

Thus, the constants A_n and ϕ_n, appearing in the general expression (5.41) for the time evolution of a uniform string with fixed ends, are ultimately determined by integrals over the string's initial displacement and velocity which are of the form (5.54) and (5.55).

As an example, suppose that the string is initially *at rest*, so that

$$v_0(x) = 0, \tag{5.60}$$

but has the initial displacement

$$y_0(x) = 2\,A \left\{ \begin{array}{ll} x/l & 0 \leq x < l/2 \\ 1 - x/l & l/2 \leq x \leq l \end{array} \right. . \tag{5.61}$$

This *triangular* pattern is zero at both ends of the string, rising *linearly* to a peak value of A halfway along the string, and is designed to mimic the initial displacement of a guitar string which is plucked at its mid-point. See Figure 5.10. A comparison of Equations (5.55) and (5.60) reveals that, in this particular example, all of the S_n coefficients are zero. Hence, from (5.58) and (5.59), $A_n = C_n$ and $\phi_n = 0$ for all n. Thus, making use of Equations (5.41), (5.54), and (5.61), the time evolution of the string is governed by

$$y(x,t) = \sum_{n=1,\infty} A_n \sin\left(n\,\pi\,\frac{x}{l}\right)\,\cos\left(n\,2\pi\,\frac{t}{\tau}\right), \tag{5.62}$$

where $\tau = 2\,l/v$ is the oscillation period of the $n = 1$ normal mode, and

$$A_n = \frac{2}{l}\int_0^{l/2} 2\,A\,\frac{x}{l}\,\sin\left(n\,\pi\,\frac{x}{l}\right)\,dx + \frac{2}{l}\int_{l/2}^{l} 2\,A\left(1 - \frac{x}{l}\right)\sin\left(n\,\pi\,\frac{x}{l}\right)\,dx. \tag{5.63}$$

The above expression transforms to

$$A_n = A\left(\frac{2}{\pi}\right)^2\left\{\int_0^{\pi/2}\theta\,\sin(n\,\theta)\,d\theta + \int_{\pi/2}^{\pi}(\pi - \theta)\,\sin(n\,\theta)\,d\theta\right\}, \tag{5.64}$$

where $\theta = \pi\,x/l$. Integration by parts yields

$$A_n = 2\,A\,\frac{\sin(n\,\pi/2)}{(n\,\pi/2)^2}. \tag{5.65}$$

Note that $A_n = 0$ whenever n is even. We conclude that the triangular initial displacement pattern (5.61) only excites normal modes with *odd* mode numbers.

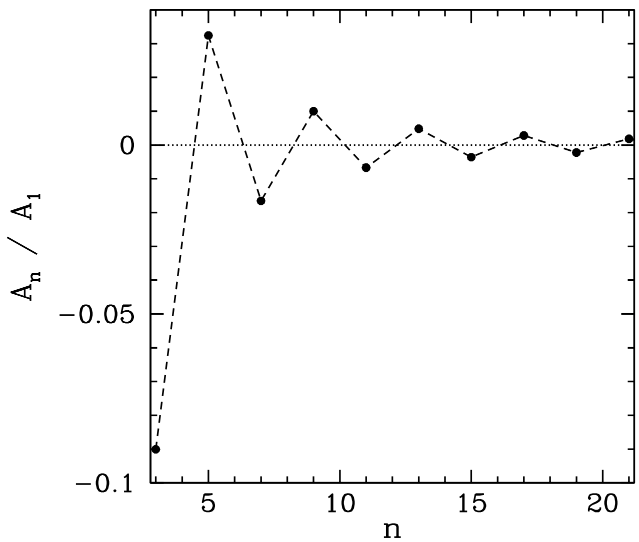


Figure 5.9: *Relative amplitudes of the overtone harmonics of a uniform guitar string plucked at its mid-point.*

Now, when a stringed instrument, such as a guitar, is sounded a characteristic pattern of normal mode oscillations is excited on the plucked string. These oscillations excite sound waves of the same frequency, which are audible to a listener. The normal mode (of appreciable amplitude) with the *lowest oscillation frequency* is called the *fundamental harmonic*, and determines the *pitch* of the musical note which is heard by the listener. For instance, a fundamental harmonic which oscillates at 261.6 Hz would correspond to "middle C". Those normal modes (of appreciable amplitude) with higher oscillation frequencies than the fundamental harmonic are called *overtone harmonics*, since their frequencies are *integer multiples* of the fundamental frequency. In general, the amplitudes of the overtone harmonics are much smaller than the amplitude of the fundamental. Nevertheless, when a stringed instrument is sounded, the particular mix of overtone harmonics which accompanies the fundamental determines the *timbre* of the musical note heard by the listener. For instance, when middle C is played on a piano and a harpsichord the same frequency fundamental harmonic is excited in both cases. However, the mix of excited overtone harmonics is quite different. This accounts for the fact that middle C played on a piano is easily distinguishable from middle C played on a harpsichord.

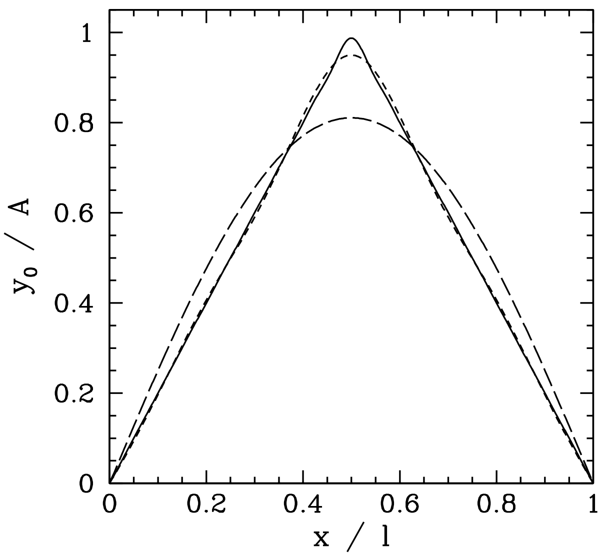


Figure 5.10: *Reconstruction of the initial displacement of a uniform guitar string plucked at its mid-point.*

Figure 5.9 shows the ratio A_n/A_1 for the first ten overtone harmonics of a uniform guitar string plucked at its midpoint: *i.e.*, the ratio A_n/A_1 for odd-n modes with $n > 1$, calculated from Equation (5.65). It can be seen that the amplitudes of the overtone harmonics are all small compared to the amplitude of the fundamental. Moreover, the amplitudes decrease rapidly in magnitude with increasing mode number, n.

In principle, we must include all of the normal modes in the sum on the right-hand side of Equation (5.62). In practice, given that the amplitudes of the normal modes decrease rapidly in magnitude as n increases, we can *truncate* the sum, by neglecting high-n normal modes, without introducing significant error into our calculation. Figure 5.11 illustrates the effect of such a truncation. In fact, this figure shows the reconstruction of $y_0(x)$, obtained by setting $t = 0$ in Equation (5.62), made with various different numbers of normal modes. The long-dashed line shows a reconstruction made with only the largest amplitude normal mode, the short dashed-line shows a reconstruction made with the four largest amplitude normal modes, and the solid line shows a reconstruction made with the sixteen largest amplitude normal modes. It can be seen that sixteen normal modes is sufficient to very accurately reconstruct the triangular initial displacement pattern. Indeed, a

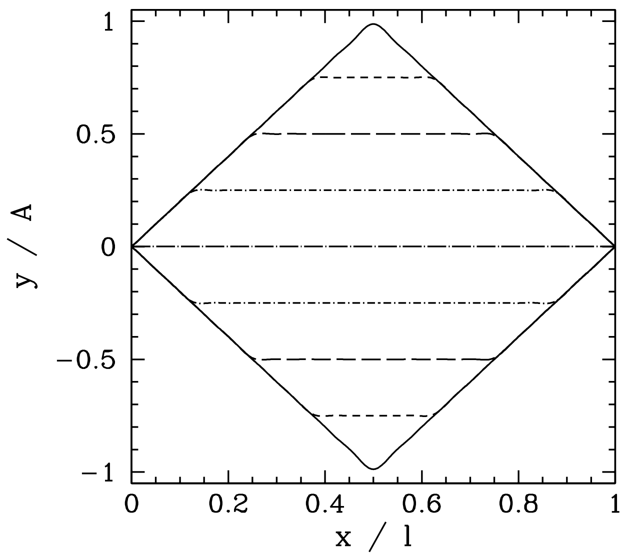


Figure 5.11: *Time evolution of a uniform guitar string plucked at its mid-point.*

reconstruction made with only four normal modes is surprisingly accurate. On the other hand, a reconstruction made with only one normal mode is fairly inaccurate.

Figure 5.11 shows the time evolution of a uniform guitar string plucked at its mid-point. This evolution is reconstructed from expression (5.62) using the sixteen largest amplitude normal modes of the string. The upper solid, upper short-dashed, upper long-dashed, upper dot-short-dashed, dot-long-dashed, lower dot-short-dashed, lower long-dashed, lower short-dashed, and lower solid curves correspond to $t/\tau = 0$, $1/16$, $1/8$, $3/16$, $1/4$, $5/16$, $3/8$, $7/16$, and $1/2$, respectively. It can be seen that the string oscillates in a fairly strange fashion. The initial kink in the string at $x = l/2$ splits into two equal kinks which propagate in opposite directions along the string at the velocity v. The string remains straight and parallel to the x-axis between the kinks, and straight and inclined to the x-axis between each kink and the closest wall. When the two kinks reach the wall the string is instantaneously found in its undisturbed position. The kinks then reflect off the two walls, with a phase change of π radians. When the two kinks meet again at $x = l/2$ the string is instantaneously found in a state which is an inverted form of its initial state. The kinks subsequently pass through one another, reflect off the walls, with another phase change of π radians, and meet for a second time at $x = l/2$. At this instant, the string is again found in its initial position. The pattern of motion then repeats itself *ad infinitum*. The period of the oscillation is the time required for a kink to propagate two string lengths, which is $\tau = 2\,l/v$. This, of course, is also the oscillation period of the $n = 1$ normal mode.

5.4 Exercises

1. Consider a uniformly beaded string with N beads which is similar to that pictured in Figure 5.1, except that each end of the string is attached to a massless ring which slides (in the y-direction) on a frictionless rod. Demonstrate that the normal modes of the system take the form

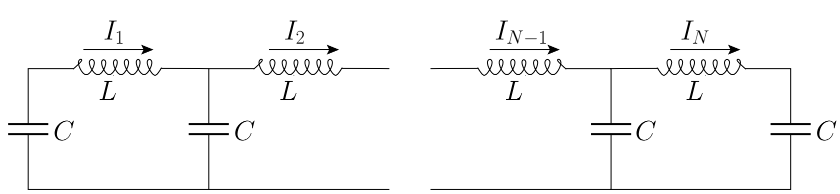
$$y_{n,i}(t) = A_n \cos\left[\frac{n\,(i-1/2)}{N}\,\pi\right]\cos(\omega_n\,t - \phi_n),$$

where

$$\omega_n = 2\,\omega_0\,\sin\left(\frac{n}{N}\,\frac{\pi}{2}\right),$$

ω_0 is as defined in Section 5.1, A_n and ϕ_n are constants, the integer $i = 1, N$ indexes the beads, and the mode number n indexes the modes. How many unique normal modes does the system possess, and what are their mode numbers? Show that the lowest frequency mode has an infinite wavelength and zero frequency. Explain this peculiar result. Plot the normal modes and normal frequencies of an $N = 8$ beaded string in a similar fashion to Figures 5.3 and 5.5.

2. Consider a uniformly beaded string with N beads which is similar to that pictured in Figure 5.1, except that the left end of the string is fixed, and the right end is attached to a massless ring which slides (in the y-direction) on a frictionless rod. Find the normal modes and normal frequencies of the system. Plot the normal modes and normal frequencies of an $N = 8$ beaded string in a similar fashion to Figures 5.3 and 5.5.



3. The above figure shows the left and right extremities of a linear LC network consisting of N identical inductors of inductance L, and $N+1$ identical capacitors of capacitance C. Let the instantaneous current flowing through the

ith inductor be $I_i(t)$, for $i = 1, N$. Demonstrate from Kirchoff's circuital laws that the currents evolve in time according to the coupled equations

$$\ddot{I}_i = \omega_0^2\,(I_{i-1} - 2\,I_i + I_{i+1}),$$

for $i = 1, N$, where $\omega_0 = 1/\sqrt{L\,C}$, and $I_0 = I_{N+1} = 0$. Find the normal frequencies of the system.

4. Suppose that the outermost two capacitors in the circuit considered in the previous exercise are short-circuited. Find the new normal frequencies of the system.

5. A uniform string of length l, tension T, and mass per unit length ρ, is stretched between two immovable walls. Suppose that the string is initially in its equilibrium state. At $t = 0$ it is struck by a hammer in such a manner as to impart an impulsive velocity u_0 to a small segment of length $a < l$ centered on the mid-point. Find an expression for the subsequent motion of the string. Plot the motion as a function of time in a similar fashion to Figure 5.11, assuming that $a = l/10$.

6. A uniform string of length l, tension T, and mass per unit length ρ, is stretched between two massless rings, attached to its ends, which slide (in the y-direction) along frictionless rods. Demonstrate that, in this case, the most general solution to the wave equation takes the form

$$y(x,t) = Y_0 + V_0\,t + \sum_{n>0} A_n\,\cos\left(n\,\pi\,\frac{x}{l}\right)\cos\left(n\,\pi\,\frac{v\,t}{l} - \phi_n\right),$$

where $v = \sqrt{T/\rho}$, and Y_0, V_0, A_n, and ϕ_n are arbitrary constants. Show that

$$\frac{2}{l}\int_0^l \cos\left(n\,\pi\,\frac{x}{l}\right)\cos\left(n'\,\pi\frac{x}{l}\right)dx = \delta_{n,n'},$$

where n and n' are integers. Use this result to demonstrate that the arbitrary constants in the above solution can be determined from the initial conditions as follows:

$$\begin{aligned} Y_0 &= \frac{2}{l}\int_0^l y_0(x)\,dx, \\ V_0 &= \frac{2}{l}\int_0^l v_0(x)\,dx, \\ A_n &= (C_n^2 + S_n^2)^{1/2}, \\ \phi_n &= \tan^{-1}(S_n/C_n), \end{aligned}$$

where $y_0(x) \equiv y(x,0)$, $v_0(x) \equiv \partial y(x,0)/\partial t$, and

$$C_n = \frac{2}{l}\int_0^l y_0(x)\,\cos\left(n\,\pi\,\frac{x}{l}\right)dx,$$

$$S_n = \frac{2}{l}\int_0^l v_0(x)\,\cos\left(n\,\pi\,\frac{x}{l}\right)\,dx.$$

Suppose that the string is initially in its equilibrium state. At $t = 0$ it is struck by a hammer in such a manner as to impart an impulsive velocity u_0 to a small segment of length $a < l$ centered on the mid-point. Find an expression for the subsequent motion of the string. Plot the motion as a function of time in a similar fashion to Figure 5.11, assuming that $a = l/10$.

7. The linear LC circuit considered in Exercise 3 can be thought of as a discrete model of a uniform lossless transmission line: *e.g.*, a co-axial cable. In this interpretation, $I_i(t)$ represents $I(x_i, t)$, where $x_i = i\,\delta x$. Moreover, $C = \mathcal{C}\,\delta x$, and $L = \mathcal{L}\,\delta x$, where $\mathcal{C}$ and $\mathcal{L}$ are the capacitance per unit length and the inductance per unit length of the line, respectively. Show that, in the limit $\delta x \to 0$, the evolution equation for the coupled currents given in Exercise 3 reduces to the wave equation

$$\frac{\partial^2 I}{\partial t^2} = v^2\,\frac{\partial^2 I}{\partial x^2},$$

where $I = I(x, t)$, x measures distance along the line, and $v = 1/\sqrt{\mathcal{L}\,\mathcal{C}}$. If $V_i(t)$ is the potential difference (measured from the top to the bottom) across the $i+1$th capacitor (from the left) in the circuit shown in Exercise 3, and $V(x, t)$ is the corresponding voltage in the transmission line, show that the discrete circuit equations relating the $I_i(t)$ and $V_i(t)$ reduce to

$$\frac{\partial V}{\partial t} = -\frac{1}{\mathcal{C}}\,\frac{\partial I}{\partial x},$$
$$\frac{\partial I}{\partial t} = -\frac{1}{\mathcal{L}}\,\frac{\partial V}{\partial x},$$

in the transmission line limit. Hence, demonstrate that the voltage in a transmission line satisfies the wave equation

$$\frac{\partial^2 V}{\partial t^2} = v^2\,\frac{\partial^2 V}{\partial x^2}.$$

8. Consider a uniform string of length l, tension T, and mass per unit length ρ which is stretched between two immovable walls. Show that the total energy of the string, which is the sum of its kinetic and potential energies, is

$$E = \frac{1}{2}\int_0^l \left[\rho\left(\frac{\partial y}{\partial t}\right)^2 + T\left(\frac{\partial y}{\partial x}\right)^2\right] dx,$$

where $y(x, t)$ is the string's (relatively small) transverse displacement. Now, the general motion of the string can be represented as a linear superposition of the normal modes:

$$y(x, t) = \sum_{n=1,\infty} A_n\,\sin\left(n\,\pi\,\frac{x}{l}\right)\cos\left(n\,\pi\,\frac{v\,t}{l} - \phi_n\right),$$

where $v = \sqrt{T/\rho}$. Demonstrate that

$$E = \sum_{n=1,\infty} E_n,$$

where

$$E_n = \frac{1}{4}\, m\, \omega_n^2\, A_n^2$$

is the energy of the nth normal mode. Here, $m = \rho\, l$ is the mass of the string, and $\omega_n = n\, \pi\, v/l$ the angular frequency of the nth normal mode.

6 Longitudinal Standing Waves

6.1 Spring Coupled Masses

Consider a mechanical system consisting of a linear array of N identical masses, m, which are free to slide in one dimension over a frictionless horizontal surface. Suppose that the masses are coupled to their immediate neighbors via identical light springs of unstretched length a, and force constant K. (Here, we use the symbol K to denote the spring force constant, rather than k, since k is already being used to denote wavenumber.) Let x measure distance along the array (from the left to the right). So, if the array is in its equilibrium configuration then the x-coordinate of the ith mass is $x_i = i\,a$, for $i = 1, N$. Consider *longitudinal* oscillations of the masses: *i.e.*, oscillations such that the x-coordinate of the ith mass becomes

$$x_i = i\,a + \psi_i(t), \tag{6.1}$$

where $\psi_i(t)$ represents the mass's longitudinal displacement from equilibrium. It is assumed that all of the displacements are relatively small: *i.e.*, $|\psi_i| \ll a$, for $i = 1, N$.

Consider the equation of motion of the ith mass. See Figure 6.1. The extensions of the springs to the immediate left and right of the mass are $\psi_i - \psi_{i-1}$ and $\psi_{i+1} - \psi_i$, respectively. Thus, the x-directed forces that these springs exert on the mass are $-K\,(\psi_i - \psi_{i-1})$ and $K\,(\psi_{i+1} - \psi_i)$, respectively, and its equation of motion is easily shown to be

$$\ddot{\psi}_i = \omega_0^{\,2}\,(\psi_{i-1} - 2\,\psi_i + \psi_{i+1}), \tag{6.2}$$

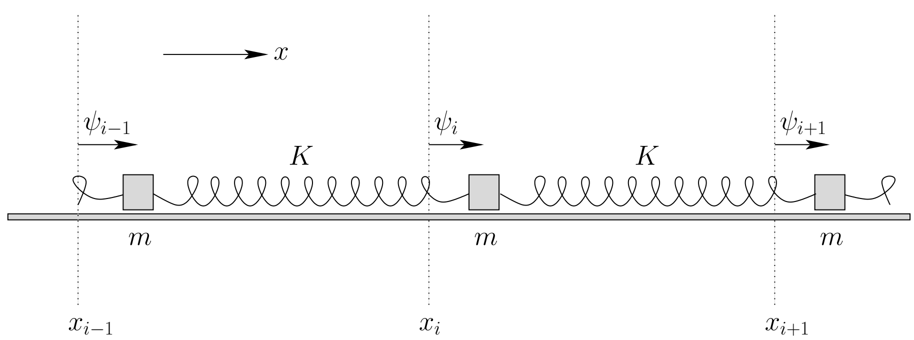


Figure 6.1: *Detail of a system of spring coupled masses.*

where $\omega_0 = \sqrt{K/m}$. Since there is nothing special about the ith mass, the above equation is assumed to hold for all N masses: *i.e.*, for $i = 1, N$. Note that Equation (6.2), which governs the *longitudinal* oscillations of a linear array of spring coupled masses, is analogous in form to Equation (5.8), which governs the *transverse* oscillations of a beaded string. This observation suggests that longitudinal and transverse waves in discrete dynamical systems (*i.e.*, systems with a finite number of degrees of freedom) can be described using the *same* mathematical equations.

We can interpret the quantities ψ_0 and ψ_{N+1}, which appear in the equations of motion for ψ_1 and ψ_N, respectively, as the longitudinal displacements of the left and right extremities of springs which are attached to the outermost masses in such a manner as to form the left and right boundaries of the array. The respective equilibrium positions of these extremities are $x_0 = 0$ and $x_{N+1} = (N+1)\,a$. Now, the end displacements, ψ_0 and ψ_{N+1}, must be *prescribed*, otherwise Equations (6.2) do not constitute a complete set of equations: *i.e.*, there are more unknowns than equations. The particular choice of ψ_0 and ψ_{N+1} depends on the nature of the physical *boundary conditions* at the two ends of the array. Suppose that the left extremity of the leftmost spring is anchored in an immovable wall. This implies that $\psi_0 = 0$: *i.e.*, the left extremity of the spring cannot move. Suppose, on the other hand, that the left extremity of the leftmost spring is not attached to anything. In this case, there is no reason for the spring to become extended, which implies that $\psi_0 = \psi_1$. In other words, if the left end of the array is *fixed* (*i.e.*, attached to an immovable object) then $\psi_0 = 0$, and if the left end is *free* (*i.e.*, not attached to anything) then $\psi_0 = \psi_1$. Likewise, if the right end of the array is fixed then $\psi_{N+1} = 0$, and if the right end is free then $\psi_{N+1} = \psi_N$.

Suppose, for the sake of argument, that the left end of the array is free, and the right end is fixed. It follows that $\psi_0 = \psi_1$, and $\psi_{N+1} = 0$. Let us search for normal modes of the general form

$$\psi_i(t) = A\,\cos[k\,(x_i - a/2)]\,\cos(\omega\,t - \phi), \tag{6.3}$$

where $A > 0$, $k > 0$, $\omega > 0$, and ϕ are constants. Note that the above expression automatically satisfies the boundary condition $\psi_0 = \psi_1$. This follows because $x_0 = 0$ and $x_1 = a$, and, consequently, $\cos[k\,(x_0 - a/2)] = \cos(-k\,a/2) = \cos(k\,a/2) = \cos[k\,(x_1 - a/2)]$. The other boundary condition, $\psi_{N+1} = 0$, is satisfied provided

$$\cos[k\,(x_{N+1} - a/2)] = \cos[(N + 1/2)\,k\,a] = 0, \tag{6.4}$$

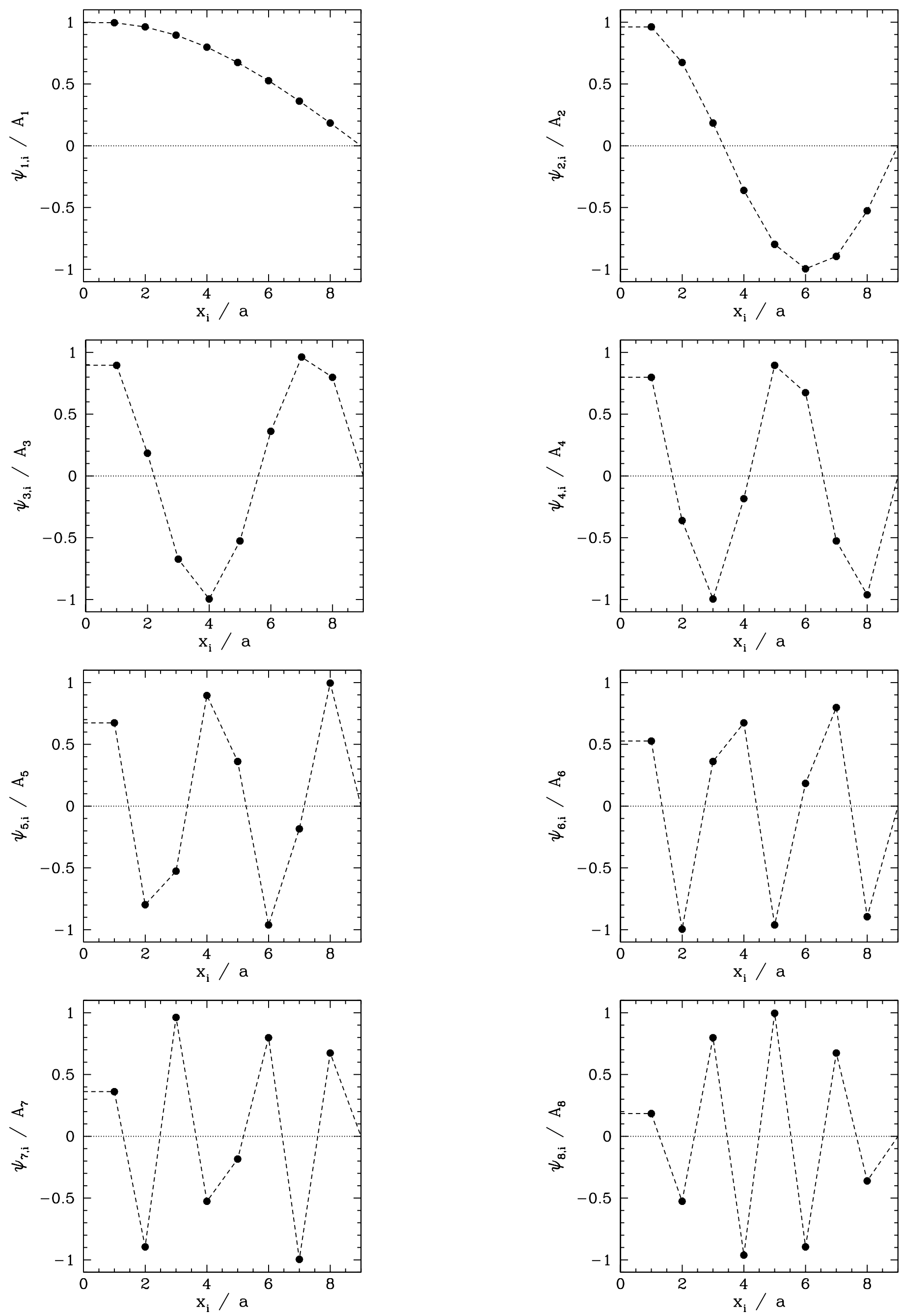


Figure 6.2: *Normal modes of a system of eight spring coupled masses.*

which yields [*cf.*, (5.15)]

$$k\,a = \frac{(n-1/2)\,\pi}{N+1/2}, \tag{6.5}$$

where n is an integer. As before, the imposition of the boundary conditions causes a quantization of the possible mode wavenumbers (see Section 5.1). Finally, substitution of (6.3) into (6.2) gives the dispersion relation [*cf.*, (5.13)]

$$\omega = 2\,\omega_0\,\sin(k\,a/2). \tag{6.6}$$

It follows, from the above analysis, that the longitudinal normal modes of a linear array of spring coupled masses, the left end of which is free, and the right end fixed, are associated with the following characteristic displacement patterns:

$$\psi_{n,i}(t) = A_n\,\cos\left[\frac{(n-1/2)\,(i-1/2)}{N+1/2}\,\pi\right]\cos(\omega_n - \phi_n), \tag{6.7}$$

where

$$\omega_n = 2\,\omega_0\,\sin\left(\frac{n-1/2}{N+1/2}\,\frac{\pi}{2}\right), \tag{6.8}$$

and the A_n and ϕ_n are arbitrary constants determined by the initial conditions. Here, the integer $i = 1, N$ indexes the masses, and the mode number n indexes the normal modes. It is easily demonstrated that there are only N unique normal modes, corresponding to mode numbers in the range 1 to N.

Figures 6.2 and 6.3 display the normal modes and normal frequencies of a linear array of eight spring coupled masses, the left end of which is free, and the right end fixed. The data shown in these figures is obtained from Equations (6.7) and (6.8), respectively, with $N = 8$. The modes in Figure 6.2 are all plotted at the instances in time at which they attain their maximum amplitudes: *i.e.*, when $\cos(\omega_n\,t - \phi_n) = 1$. It can be seen that normal modes with small wavenumbers—*i.e.*, $k\,a \ll 1$, so that $n \ll N$—have displacements which vary in a fairly smooth sinusoidal manner from mass to mass, and oscillations frequencies which increase approximately linearly with increasing wavenumber. On the other hand, normal modes with large wavenumbers—*i.e.*, $k\,a \sim 1$, so that $n \sim N$—have displacements which exhibit large variations from mass to mass, and oscillation frequencies which do not depend linearly on wavenumber. We conclude that the longitudinal normal modes of an array of spring coupled masses have analogous properties to the transverse normal modes of a beaded string. See Section 5.1.

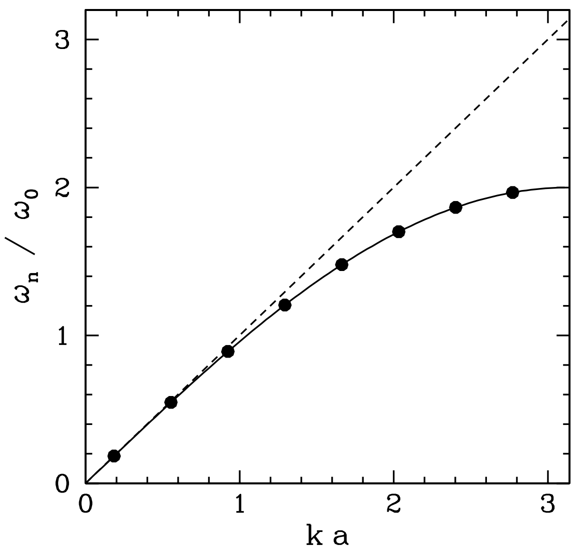


Figure 6.3: *Normal frequencies of a system of eight spring coupled masses.*

The dynamical system pictured in Figure 6.1 can be used to model the effect of a planar *sound wave* (*i.e.*, a longitudinal oscillation in position which is periodic in space in one dimension) on a *crystal lattice*. In this application, the masses represent parallel planes of atoms, the springs represent the interatomic forces acting between these planes, and the longitudinal oscillations represent the sound wave. Of course, a macroscopic crystal contains a great many atomic planes, so we would expect N to be very large. Note, however, from Equations (6.5) and (6.8), that, no matter how large N becomes, $k\,a$ cannot exceed π (since n cannot exceed N), and ω_n cannot exceed $2\,\omega_0$. In other words, there is a *minimum wavelength* that a sound wave in a crystal lattice can have, which turns out to be twice the interatomic spacing, and a corresponding *maximum oscillation frequency*. For waves whose wavelengths are much greater than the interatomic spacing (*i.e.*, $k\,a \ll 1$), the dispersion relation (6.6) reduces to

$$\omega \simeq k\,c \tag{6.9}$$

where $c = \omega_0\,a = \sqrt{K/m}\,a$ is a constant which has the dimensions of velocity. It seems plausible that (6.9) is the dispersion relation for sound waves in a *continuous* elastic medium. Let us investigate such waves.

6.2 Sound Waves in an Elastic Solid

Consider a thin uniform elastic rod of length l and cross-sectional area A. Let us examine the longitudinal oscillations of such a rod. These oscillations are usually, somewhat loosely, referred to as *sound waves*. It is again convenient to let x denote position along the rod. Thus, in equilibrium, the two ends of the rod lie at $x = 0$ and $x = l$. Suppose that a sound wave causes an x-directed displacement $\psi(x,t)$ of the various elements of the rod from their equilibrium positions. Consider a thin section of the rod, of length δx, lying between $x - \delta x/2$ and $x + \delta x/2$. The displacements of the left and right boundaries of the section are $\psi(x - \delta x/2, t)$ and $\psi(x + \delta x/2, t)$, respectively. Thus, the change in length of the section, due to the action of the sound wave, is $\psi(x + \delta x/2, t) - \psi(x - \delta x/2, t)$. Now, *strain* in an elastic rod is defined as *change in length over unperturbed length*. Thus, the strain in the section of the rod under consideration is

$$\epsilon(x,t) = \frac{\psi(x + \delta x/2, t) - \psi(x - \delta x/2, t)}{\delta x}. \tag{6.10}$$

In the limit $\delta x \to 0$, this becomes

$$\epsilon(x,t) = \frac{\partial \psi(x,t)}{\partial x}. \tag{6.11}$$

Of course, it is assumed that the strain is small: *i.e.*, $|\epsilon| \ll 1$. *Stress*, $\sigma(x,t)$, in an elastic rod is defined as the *elastic force per unit cross-sectional area*. In a conventional elastic material, the relationship between stress and strain (for small strains) takes the simple form

$$\sigma = Y\,\epsilon. \tag{6.12}$$

Here, Y is a constant, with the dimensions of pressure, which is known as the *Young's modulus*. Note that if the strain in a given element is positive then the stress acts to lengthen the element, and *vice versa*. (Similarly, in the spring coupled mass system investigated in the previous section, the external forces exerted on an individual spring act to lengthen it when its extension is positive, and *vice versa*.)

Consider the motion of a thin section of the rod lying between $x - \delta x/2$ and $x + \delta x/2$. If ρ is the mass density of the rod then the section's mass is $\rho\,A\,\delta x$. The stress acting on the left boundary of the section is $\sigma(x - \delta x/2) = Y\,\epsilon(x - \delta x/2)$. Since stress is force per unit area, the force acting on the left boundary is $A\,Y\,\epsilon(x - \delta x/2)$. This force is directed in the minus x-direction, assuming that the strain is positive (*i.e.*, the force acts to lengthen the section). Likewise, the force acting on the right boundary of the section is $A\,Y\,\epsilon(x+\delta x/2)$, and is directed in the positive x-direction, assuming that the strain is positive (*i.e.*, the force again acts to lengthen the section). Finally, the mean longitudinal (*i.e.*, x-directed) acceleration of the section is $\partial^2\psi(x,t)/\partial t^2$. Hence, the section's longitudinal equation of motion becomes

$$\rho\,A\,\delta x\,\frac{\partial^2\psi(x,t)}{\partial t^2} = A\,Y\,[\epsilon(x+\delta x/2,t) - \epsilon(x-\delta x/2)]\,. \tag{6.13}$$

In the limit $\delta x \to 0$, this expression reduces to

$$\rho\,\frac{\partial^2\psi(x,t)}{\partial t^2} = Y\,\frac{\partial\epsilon(x,t)}{\partial x}, \tag{6.14}$$

or

$$\frac{\partial^2\psi}{\partial t^2} = c^2\,\frac{\partial^2\psi}{\partial x^2}, \tag{6.15}$$

where $c = \sqrt{Y/\rho}$ is a constant having the dimensions of velocity, which turns out to be the speed of sound in the rod (see Section 7.1), and use has been made of Equation (6.11). Of course, (6.15) is a *wave equation*. As such, it has the same form as Equation (5.27), which governs the motion of transverse waves on a uniform string. This suggests that longitudinal and transverse waves in continuous dynamical systems (*i.e.*, systems with an infinite number of degrees of freedom) can be described using the *same* mathematical equations.

In order to solve (6.15), we need to specify *boundary conditions* at the two ends of the rod. Suppose that the left end of the rod is *fixed*: *i.e.*, it is clamped in place so that it cannot move. This implies that $\psi(0,t) = 0$. Suppose, on the other hand, that the left end of the rod is *free*: *i.e.*, it is not attached to anything. This implies that $\sigma(0,t) = 0$, since there is nothing that the end can exert a force (or a stress) on, and *vice versa*. It follows from (6.11) and (6.12) that $\partial\psi(0,t)/\partial x = 0$. Likewise, if the right end of the rod is fixed then $\psi(l,t) = 0$, and if the right end is free then $\partial\psi(l,t)/\partial x = 0$.

Suppose, for the sake of argument, that the left end of the rod is free, and the right end is fixed. It follows that $\partial\psi(0,t)/\partial x = 0$, and $\psi(l,t) = 0$. Let us search for normal modes of the form

$$\psi(x,t) = A\,\cos(k\,x)\,\cos(\omega\,t - \phi), \tag{6.16}$$

where $A > 0$, $k > 0$, $\omega > 0$, and ϕ are constants. Note that the above expression automatically satisfies the boundary condition $\partial\psi(0,t)/\partial x = 0$.

The other boundary condition is satisfied provided

$$\cos(k\,l) = 0, \tag{6.17}$$

which yields

$$k\,l = (n - 1/2)\,\pi, \tag{6.18}$$

where n is an integer. As usual, the imposition of the boundary conditions leads to a quantization of the possible mode wavenumbers. Substitution of (6.16) into the equation of motion (6.15) yields the normal mode dispersion relation

$$\omega = k\,c = k\sqrt{\frac{Y}{\rho}}. \tag{6.19}$$

Note that this dispersion relation is consistent with the previously derived dispersion relation (6.9), since $m = \rho\,A\,a$ and $K = A\,Y/a$. Here, a is the interatomic spacing, m the mass of a section of the rod containing a single plane of atoms, and K the effective force constant between neighboring atomic planes.

It follows, from the above analysis, that the nth longitudinal normal mode of an elastic rod, of length l, whose left end is free, and whose right end is fixed, is associated with the characteristic displacement pattern

$$\psi_n(x, t) = A_n \cos\left[(n - 1/2)\,\pi\,\frac{x}{l}\right]\cos(\omega_n\,t - \phi_n), \tag{6.20}$$

where

$$\omega_n = (n - 1/2)\,\frac{\pi\,c}{l}. \tag{6.21}$$

Here, A_n and ϕ_n are constants which are determined by the initial conditions. It is easily demonstrated that only those normal modes whose mode numbers are *positive integers* yield unique displacement patterns: *i.e.*, $n > 0$. Equation (6.20) describes a *standing wave* whose nodes (*i.e.*, points at which $\psi = 0$ for all t) are evenly spaced a distance $l/(n - 1/2)$ apart. Of course, the boundary condition $\psi(l, t) = 0$ ensures that the right end of the rod is always coincident with a node. On the other hand, the boundary condition $\partial\psi(0, t)/\partial x = 0$ ensures that the left hand of the rod is always coincident with a point of maximum amplitude oscillation [*i.e.*, a point at which $\cos(k\,x) = \pm 1$]. Such a point is known as an *anti-node*. It is easily demonstrated that the anti-nodes associated with a given normal mode lie *halfway* between the corresponding nodes. Note, from (6.21), that the normal mode oscillation frequencies depend *linearly* on mode number. Finally, it is easily demonstrated that, in the long wavelength limit $k\,a \ll 1$, the normal modes and normal frequencies of a uniform elastic rod specified in Equations (6.20) and (6.21) are analagous to the normal modes and normal frequencies of a linear array of identical spring coupled masses specified in Equations (6.7) and (6.8), and pictured in Figures 6.2 and 6.3.

Since Equation (6.15) is obviously linear, its most general solution is a *linear combination* of all of the normal modes: *i.e.*,

$$\psi(x,t) = \sum_{n'=1,\infty} A_{n'}\,\cos\left[(n'-1/2)\,\pi\,\frac{x}{l}\right]\,\cos\left[(n'-1/2)\,\pi\,\frac{c\,t}{l}-\phi_{n'}\right]. \tag{6.22}$$

The constants A_n and ϕ_n are determined from the initial displacement,

$$\psi(x,0) = \sum_{n'=1,\infty} A_{n'}\,\cos\phi_{n'}\,\cos\left[(n'-1/2)\,\pi\,\frac{x}{l}\right], \tag{6.23}$$

and the initial velocity,

$$\dot{\psi}(x,0) = \frac{\pi\,c}{l}\sum_{n'=1,\infty}(n'-1/2)\,A_{n'}\,\sin\phi_{n'}\,\cos\left[(n'-1/2)\,\pi\,\frac{x}{l}\right]. \tag{6.24}$$

Now, it is easily demonstrated that [*cf.*, (5.50)]

$$\frac{2}{l}\int_0^l \cos\left[(n-1/2)\,\pi\,\frac{x}{l}\right]\,\cos\left[(n'-1/2)\,\pi\,\frac{x}{l}\right]dx = \delta_{n,n'}. \tag{6.25}$$

Thus, multiplying (6.23) by $(2/l)\,\cos[(n-1/2)\,\pi\,x/l]$, and then integrating over x from 0 to l, we obtain

$$C_n \equiv \frac{2}{l}\int_0^l \psi(x,0)\,\cos\left[(n-1/2)\,\pi\,\frac{x}{l}\right]dx = A_n\,\cos\phi_n, \tag{6.26}$$

where use has been made of (6.25) and (5.51). Likewise, (6.24) gives

$$S_n \equiv \frac{2}{c\,(n-1/2)\,\pi}\int_0^l \dot{\psi}(x,0)\,\cos\left[(n-1/2)\,\pi\,\frac{x}{l}\right]dx = A_n\,\sin\phi_n. \tag{6.27}$$

Finally, $A_n = (C_n^{\,2}+S_n^{\,2})^{1/2}$ and $\phi_n = \tan^{-1}(S_n/C_n)$.

Suppose, for the sake of example, that the rod is initially at rest, and that its left end is hit with a hammer at $t=0$ in such a manner that a section of the rod lying between $x=0$ and $x=a$ (where $a<l$) acquires an instantaneous velocity V_0. It follows that $\psi(0,t)=0$. Furthermore, $\dot{\psi}(0,t)=V_0$ if

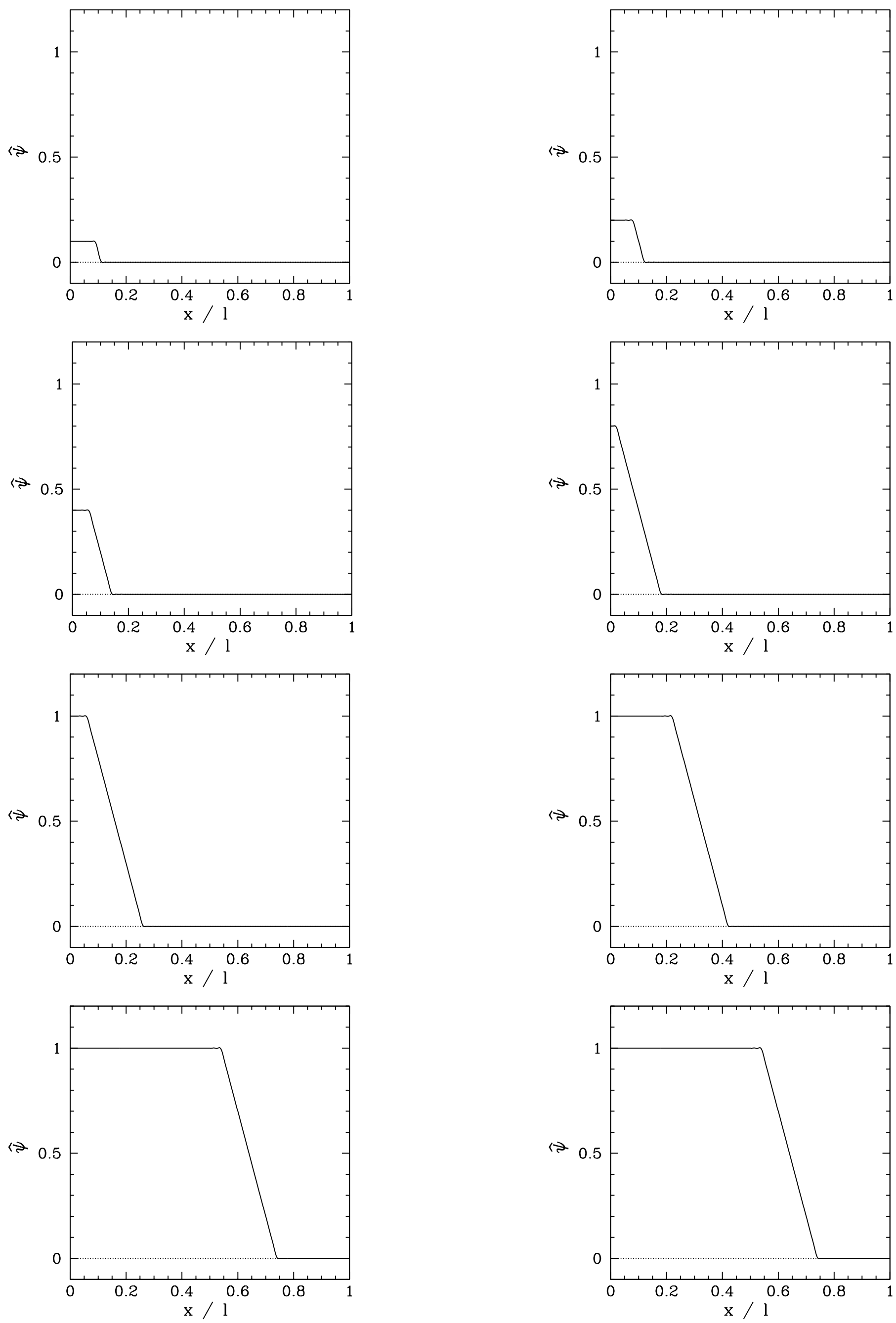


Figure 6.4: *Time evolution of the normalized displacement of an elastic rod.*

$0 \leq x \leq a$, and $\dot{\psi}(0, t) = 0$ otherwise. It is easily demonstrated that these initial conditions yield $C_n = 0$, $\phi_n = \pi/2$,

$$A_n = S_n = \frac{V_0\,a}{c}\,\frac{2}{\pi}\,\frac{\sin[(n-1/2)\,\pi\,a/l]}{(n-1/2)^2\,\pi\,a/l}, \tag{6.28}$$

and

$$\psi(x,t) = \sum_{n=1,\infty} A_n \cos\left[(n-1/2)\,\pi\,\frac{x}{l}\right] \sin\left[(n-1/2)\,\pi\,\frac{t}{\tau}\right], \tag{6.29}$$

where $\tau = l/c$. Figure 6.4 shows the time evolution of the normalized rod displacement, $\hat{\psi}(x,t) = (c/V_0\,a)\,\psi(x,t)$, calculated from the above equations using the first 100 normal modes (*i.e.*, $n = 1, 100$), and choosing $a/l = 0.1$. The top-left, top-right, middle-left, middle-right, bottom-left, and bottom-right panels correspond to $t/\tau = 0.01$, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, and 1.28, respectively. It can be seen that the hammer blow generates a displacement wave that initially develops at the free end of the rod ($x/l = 0$), which is the end that is struck, propagates along the rod at the velocity c, and reflects off the fixed end ($x/l = 1$) at time $t/\tau = 1$ with no phase shift.

6.3 Sound Waves in an Ideal Gas

Consider a uniform ideal gas of equilibrium mass density ρ and equilibrium pressure p. Let us investigate the longitudinal oscillations of such a gas. Of course, these oscillations are usually referred to as *sound waves*. Generally speaking, a sound wave in an ideal gas oscillates sufficiently rapidly that heat is unable to flow fast enough to smooth out any temperature perturbations generated by the wave. Under these circumstances, the gas obeys the *adiabatic gas law*,

$$p\,V^\gamma = \text{constant}, \tag{6.30}$$

where p is the pressure, V the volume, and γ the *ratio of specific heats* (*i.e.*, the ratio of the gas's specific heat at constant pressure to its specific heat at constant volume). This ratio is approximately 1.4 for ordinary air.

Consider a sound wave in a column of gas of cross-sectional area A. Let x measure distance along the column. Suppose that the wave generates an x-directed displacement of the column, $\psi(x,t)$. Consider a small section of the column lying between $x - \delta x/2$ and $x + \delta x/2$. The change in volume of the section is $\delta V = A\,[\psi(x+\delta x/2,t) - \psi(x-\delta x/2,t)]$. Hence, the relative change in volume, which is assumed to be small, is

$$\frac{\delta V}{V} = \frac{A\,[\psi(x+\delta x/2,t) - \psi(x-\delta x/2,t)]}{A\,\delta x}. \tag{6.31}$$

In the limit $\delta x \to 0$, this becomes

$$\frac{\delta V(x,t)}{V} = \frac{\partial\psi(x,t)}{\partial x}. \tag{6.32}$$

The pressure perturbation $\delta p(x,t)$ associated with the volume perturbation $\delta V(x,t)$ follows from (6.30), which yields

$$(p+\delta p)\,(V+\delta V)^{\gamma} = p\,V^{\gamma}, \tag{6.33}$$

or

$$(1+\delta p/p)\,(1+\delta V/V)^{\gamma} \simeq 1 + \delta p/p + \gamma\,\delta V/V = 1, \tag{6.34}$$

giving

$$\delta p = -\gamma\,p\,\frac{\delta V}{V} = -\gamma\,p\,\frac{\partial\psi}{\partial x}, \tag{6.35}$$

where use has been made of (6.32).

Consider a section of the gas column lying between $x - \delta x/2$ and $x + \delta x/2$. The mass of this section is $\rho\,A\,\delta x$. The x-directed force acting on its left boundary is $A\,[p + \delta p(x-\delta x/2,t)]$, whereas the x-directed force acting on its right boundary is $-A\,[p + \delta p(x+\delta x/2,t)]$. Finally, the average longitudinal (*i.e.*, x-directed) acceleration of the section is $\partial^2\psi(x,t)/\partial t^2$. Thus, the section's longitudinal equation of motion is written

$$\rho\,A\,\delta x\,\frac{\partial^2\psi(x,t)}{\partial t^2} = -A\,[\delta p(x+\delta x/2,t) - \delta p(x-\delta x/2,t)]\,. \tag{6.36}$$

In the limit $\delta x \to 0$, this equation reduces to

$$\rho\,\frac{\partial^2\psi(x,t)}{\partial t^2} = -\frac{\partial\delta p(x,t)}{\partial x}. \tag{6.37}$$

Finally, (6.35) yields

$$\frac{\partial^2\psi}{\partial t^2} = c^2\,\frac{\partial^2\psi}{\partial x^2}, \tag{6.38}$$

where $c = \sqrt{\gamma\,p/\rho}$ is a constant with the dimensions of velocity, which turns out to be the sound speed in the gas (see Section 7.1).

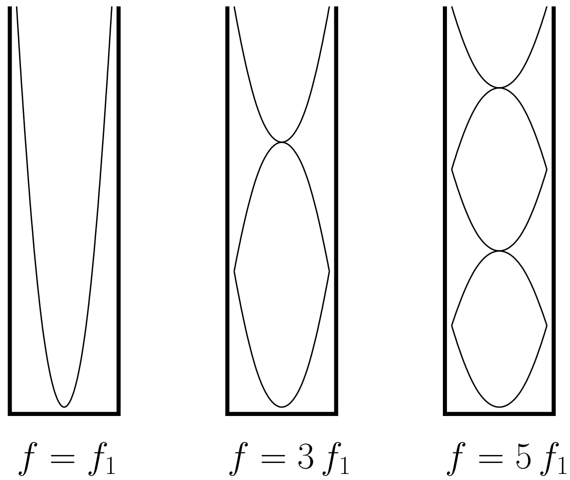


Figure 6.5: *First three normal modes of an organ pipe.*

As an example, suppose that a standing wave is excited in a uniform organ pipe of length l. Let the closed end of the pipe lie at $x = 0$, and the open end at $x = l$. The standing wave satisfies the wave equation (6.38), where c represents the speed of sound in air. The boundary conditions are that $\psi(0,t) = 0$—*i.e.*, there is zero longitudinal displacement of the air at the closed end of the pipe—and $\partial\psi(l,t)/\partial x = 0$—*i.e.*, there is zero pressure perturbation at the open end of the pipe (since the small pressure perturbation in the pipe is not intense enough to modify the pressure of the air external to the pipe). Let us write the displacement pattern associated with the standing wave in the form

$$\psi(x,t) = A\,\sin(k\,x)\,\cos(\omega\,t - \phi), \tag{6.39}$$

where $A > 0$, $k > 0$, $\omega > 0$, and ϕ are constants. This expression automatically satisfies the boundary condition $\psi(0,t) = 0$. The other boundary condition is satisfied provided

$$\cos(k\,l) = 0, \tag{6.40}$$

which yields

$$k\,l = (n - 1/2)\,\pi, \tag{6.41}$$

where the mode number n is a positive integer. Equations (6.38) and (6.39) yield the dispersion relation

$$\omega = k\,c. \tag{6.42}$$

Hence, the nth normal mode has a wavelength

$$\lambda_n = \frac{4\,l}{2n-1}, \tag{6.43}$$

and an oscillation frequency (in Hertz)

$$f_n = (2n-1)\,f_1, \tag{6.44}$$

where $f_1 = c/4\,l$ is the frequency of the fundamental harmonic (*i.e.*, the normal mode with the lowest oscillation frequency). Figure 6.5 shows the characteristic displacement patterns (which are pictured as transverse displacements, for the sake of clarity) and oscillation frequencies of the pipe's first three normal modes (*i.e.*, $n = 1, 2$, and 3). It can be seen that the modes all have a node at the closed end of the pipe, and an anti-node at the open end. The fundamental harmonic has a wavelength which is four times the length of the pipe. The first overtone harmonic has a wavelength which is 4/3rds the length of the pipe, and a frequency which is three times that of the fundamental. Finally, the second overtone has a wavelength which is 4/5ths the length of the pipe, and a frequency which is five times that of the fundamental. By contrast, the normal modes of a guitar string have nodes at either end of the string. See Figure 5.6. Thus, as is easily demonstrated, the fundamental harmonic has a wavelength which is twice the length of the string. The first overtone harmonic has a wavelength which is the length of the string, and a frequency which is twice that of the fundamental. Finally, the second overtone harmonic has a wavelength which is 2/3rds the length of the string, and a frequency which is three times that of the fundamental.

6.4 Fourier Analysis

Playing a musical instrument, such as a guitar or an organ, generates a set of standing waves which cause a sympathetic oscillation in the surrounding air. Such an oscillation consists of a fundamental harmonic, whose frequency determines the pitch of the musical note heard by the listener, accompanied by a set of overtone harmonics which determine the timbre of the note. By definition, the oscillation frequencies of the overtone harmonics are *integer multiples* of that of the fundamental. Thus, we expect the pressure perturbation generated in a listener's ear when a musical instrument is played to have the general form

$$\delta p(t) = \sum_{n=1,\infty} A_n \cos(n\,\omega\,t - \phi_n), \tag{6.45}$$

where ω is the angular frequency of the fundamental (*i.e.*, $n = 1$) harmonic, and the A_n and ϕ_n are the amplitudes and phases of the various harmonics. The above expression can also be written

$$\delta p(t) = \sum_{n=1,\infty} \left[C_n \cos(n\,\omega\,t) + S_n \sin(n\,\omega\,t)\right], \tag{6.46}$$

where $C_n = A_n \cos\phi_n$ and $S_n = A_n \sin\phi_n$. Note that $\delta p(t)$ is *periodic in time* with period $\tau = 2\pi/\omega$. In other words, $\delta p(t+\tau) = \delta p(t)$ for all t. This follows because $\cos(\theta + n\,2\pi) = \cos\theta$ and $\sin(\theta + n\,2\pi) = \sin\theta$ for all angles, θ, and for all integers, n. [Moreover, there is no $\tau' < \tau$ for which $\delta p(t+\tau') = \delta p(t)$ for all t.] So, the question arises, can any periodic waveform be represented as a linear superposition of sine and cosine waveforms, whose periods are integer subdivisions of that of the waveform, such as that shown in Equation (6.46)? To put it another way, given an arbitrary periodic waveform $\delta p(t)$, can we uniquely determine the constants C_n and S_n appearing in expression (6.46)? Actually, it turns out that we can. Incidentally, the decomposition of a periodic waveform into a linear superposition of sinusoidal waveforms is commonly known as *Fourier analysis*. Let examine this topic in a little more detail.

The problem under investigation is as follows. Given a periodic waveform $y(t)$, where $y(t+\tau) = y(t)$ for all t, we need to determine the constants C_n and S_n in the expansion

$$y(t) = \sum_{n'=1,\infty} \left[C_{n'} \cos(n'\,\omega\,t) + S_{n'} \sin(n'\,\omega\,t)\right], \tag{6.47}$$

where $\omega = 2\pi/\tau$. Now, it is easily demonstrated that [*cf.*, (5.50)]

$$\frac{2}{\tau}\int_0^\tau \cos(n\,\omega\,t)\,\cos(n'\,\omega\,t)\,dt = \delta_{n,n'}, \tag{6.48}$$

$$\frac{2}{\tau}\int_0^\tau \sin(n\,\omega\,t)\,\sin(n'\,\omega\,t)\,dt = \delta_{n,n'}, \tag{6.49}$$

$$\frac{2}{\tau}\int_0^\tau \cos(n\,\omega\,t)\,\sin(n'\,\omega\,t)\,dt = 0, \tag{6.50}$$

where n and n' are positive integers. Thus, multiplying Equation (6.47) by $(2/\tau)\,\cos(n\,\omega\,t)$, and then integrating over t from 0 to τ, we obtain

$$C_n = \frac{2}{\tau}\int_0^\tau y(t)\,\cos(n\,\omega\,t)\,dt, \tag{6.51}$$

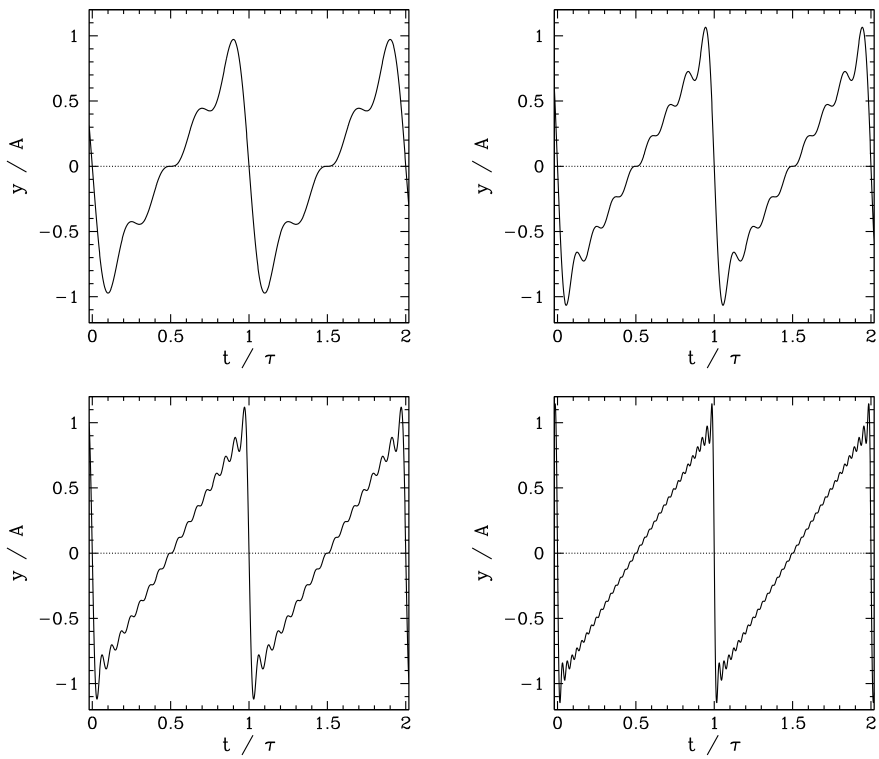


Figure 6.6: *Fourier reconstruction of a periodic sawtooth waveform.*

where use has been made of (6.48)–(6.50), as well as (5.51). Likewise, multiplying (6.47) by $(2/\tau)\,\sin(n\,\omega\,t)$, and then integrating over t from 0 to τ, we obtain

$$S_n = \frac{2}{\tau}\int_0^\tau y(t)\,\sin(n\,\omega\,t)\,dt. \tag{6.52}$$

Thus, we have uniquely determined the constants C_n and S_n in the expansion (6.47). These constants are generally known as *Fourier coefficients*, whereas the expansion itself is known as either a *Fourier expansion* or a *Fourier series*.

In principle, there is no restriction on the waveform $y(t)$ in the above analysis, other than the requirement that it be periodic in time. In other words, we ought to be able to Fourier analyze any periodic waveform. Let us see how this works. Consider the periodic sawtooth waveform (see Figure 6.6)

$$y(t) = A\,(2\,t/\tau - 1) \qquad 0 \leq t/\tau \leq 1, \tag{6.53}$$

with $y(t+\tau) = y(t)$ for all t. This waveform rises linearly from an initial value $-A$ at $t = 0$ to a final value $+A$ at $t = \tau$, discontinuously jumps back to its initial value, and then repeats *ad infinitum*. According to Equations (6.51) and (6.52), the Fourier harmonics of the waveform are

$$C_n = \frac{2}{\tau}\int_0^\tau A\,(2\,t/\tau - 1)\,\cos(n\,\omega\,t)\,dt = \frac{A}{\pi^2}\int_0^{2\pi}(\theta-\pi)\,\cos(n\,\theta)\,d\theta, \tag{6.54}$$

$$S_n = \frac{2}{\tau}\int_0^\tau A\,(2\,t/\tau - 1)\,\sin(n\,\omega\,t)\,dt = \frac{A}{\pi^2}\int_0^{2\pi}(\theta-\pi)\,\sin(n\,\theta)\,d\theta, \tag{6.55}$$

where $\theta = \omega\,t$. Integration by parts yields

$$C_n = 0, \tag{6.56}$$

$$S_n = -\frac{2\,A}{n\,\pi}. \tag{6.57}$$

Hence, the Fourier reconstruction of the waveform is written

$$y(t) = -\frac{2\,A}{\pi}\sum_{n=1,\infty}\frac{\sin(n\,2\pi\,t/\tau)}{n}. \tag{6.58}$$

Given that the Fourier coefficients fall off like $1/n$, as n increases, it seems plausible that the above series can be *truncated*, after a finite number of terms, without unduly affecting the reconstructed waveform. Figure 6.6 shows the result of truncating the series after 4, 8, 16, and 32 terms (these cases correspond the top-left, top-right, bottom-left, and bottom-right panels, respectively). It can be seen that the reconstruction becomes increasingly accurate as the number of terms retained in the series increases. The annoying oscillations in the reconstructed waveform at $t = 0$, τ, and 2τ are known as *Gibbs phenomena*, and are the inevitable consequence of trying to represent a *discontinuous* waveform as a Fourier series. In fact, it can be demonstrated mathematically that, no matter how many terms are retained in the series, the Gibbs phenomena never entirely go away.

We can slightly generalize the Fourier series (6.47) by including an $n = 0$ term. In other words,

$$y(t) = C_0 + \sum_{n'=1,\infty}\left[C_{n'}\,\cos(n'\,\omega\,t) + S_{n'}\,\sin(n'\,\omega\,t)\right], \tag{6.59}$$

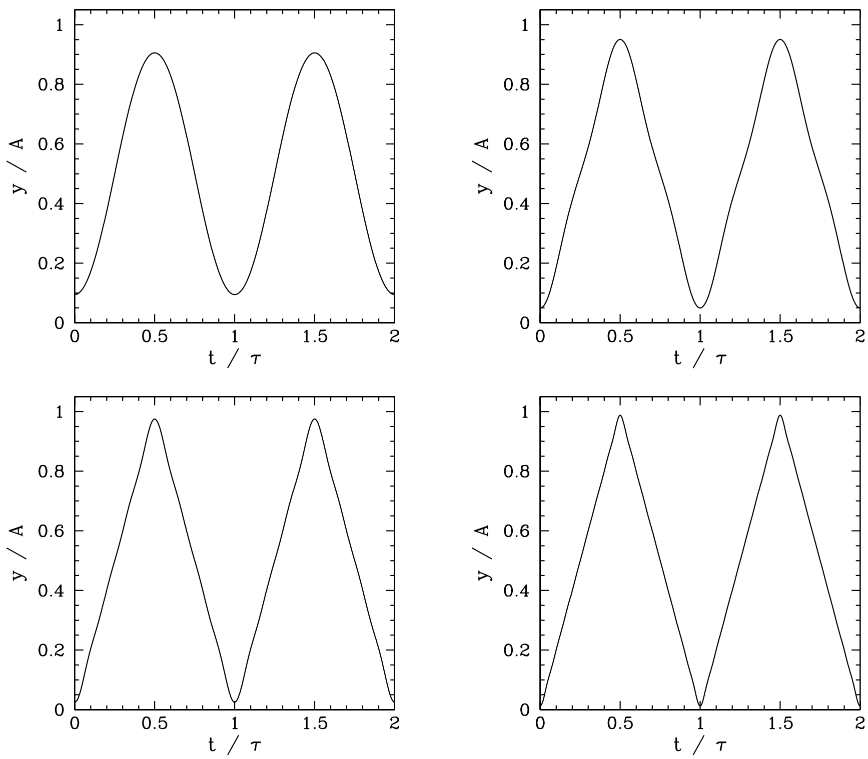


Figure 6.7: *Fourier reconstruction of a periodic "tent" waveform.*

which allows the waveform to have a non-zero average. Of course, there is no term involving S_0, since $\sin(n\,\omega\,t) = 0$ when $n = 0$. Now, it is easily demonstrated that

$$\frac{2}{\tau}\int_0^\tau \cos(n\,\omega\,t)\,dt = 0, \tag{6.60}$$

$$\frac{2}{\tau}\int_0^\tau \sin(n\,\omega\,t)\,dt = 0, \tag{6.61}$$

where $\omega = 2\pi/\tau$, and n is a positive integer. Thus, making use of these expressions, as well as Equations (6.48)–(6.50), we can easily show that

$$C_0 = \frac{1}{\tau}\int y(t)\,dt, \tag{6.62}$$

and that Equations (6.51) and (6.52) still hold for $n > 0$.

As an example, consider the periodic "tent" waveform (see Figure 6.7)

$$y(t) = 2\,A \left\{ \begin{array}{lcl} t/\tau & & 0 \leq t/\tau \leq 1/2 \\ 1 - t/\tau & & 1/2 < t/\tau \leq 1 \end{array} \right. , \tag{6.63}$$

where $y(t + \tau) = y(t)$ for all t. This waveform rises linearly from zero at $t = 0$, reaches a peak value A at $t = \tau/2$, falls linearly, becomes zero again at $t = \tau$, and repeats *ad infinitum*. Moreover, the waveform clearly has a non-zero average. It is easily demonstrated, from Equations (6.51), (6.52), (6.62), and (6.63), that

$$C_0 = \frac{A}{2}, \tag{6.64}$$

and

$$C_n = -A\,\frac{\sin^2(n\,\pi/2)}{(n\,\pi/2)^2} \tag{6.65}$$

for $n > 1$, with $S_n = 0$ for $n > 1$. Note that only the odd-n Fourier harmonics are non-zero. Figure 6.7 shows a Fourier reconstruction of the "tent" waveform using the first 1, 2, 4, and 8 terms (in addition to the C_0 term) in the Fourier series (these cases correspond to the top-left, top-right, bottom-left, and bottom-right panels, respectively). It can be seen that the reconstruction becomes increasingly accurate as the number of terms in the series increases. Moreover, in this example, there is no sign of Gibbs phenomena, since the tent waveform is completely continuous.

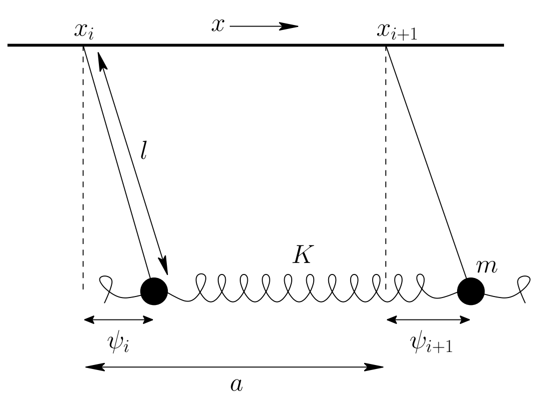
Now, in our first example—*i.e.*, the sawtooth waveform—all of the S_n Fourier coefficients are zero, whereas in our second example—*i.e.*, the tent waveform—all of the C_n coefficients are zero. It is easily demonstrated that this occurs because the sawtooth waveform is *odd* in t—*i.e.*, $y(-t) = -y(t)$ for all t—whereas the tent waveform is *even*—*i.e.*, $y(-t) = y(t)$ for all t. In fact, it is a general rule that waveforms which are even in t only have cosines in their Fourier series, whereas waveforms which are odd only have sines. Of course, waveforms which are neither even nor odd in t have both cosines and sines in their Fourier series.

Fourier series arise quite naturally in the theory of standing waves, since the normal modes of oscillation of any uniform continuous system possessing linear equations of motion (*e.g.*, a uniform string, an elastic solid, an ideal gas) take the form of spatial cosine and sine waves whose wavelengths are rational fractions of one another. Thus, the instantaneous spatial waveform of such a system can always be represented as a linear superposition of cosine and sine waves: *i.e.*, a Fourier series in space, rather than in time. In fact, we can easily appreciate that the process of determining the amplitudes and phases of the normal modes of oscillation from the initial conditions is essentially equivalent to Fourier analyzing the initial conditions in space—see Sections 5.3 and 6.2.

6.5 Exercises

1. Estimate the highest possible frequency (in Hertz), and the smallest possible wavelength, of a sound wave in aluminium, due to the discrete atomic structure of this material. The mass density, Young's modulus, and atomic weight of aluminium are $2.7 \times 10^3\,\mathrm{kg\,m^{-3}}$, $6 \times 10^{10}\,\mathrm{N\,m^{-2}}$, and 27, respectively.

2. Consider a linear array of N identical simple pendulums of mass m and length l which are suspended from equal height points that are evenly spaced a distance a apart. Suppose that each pendulum bob is attached to its two immediate neighbors by means of light springs of unstretched length a and spring constant K. The figure shows a small part of such an array. Let $x_i = i\,a$ be the equilibrium position of the ith bob, for $i = 1, N$, and let $\psi_i(t)$ be its horizontal displacement. It is assumed that $|\psi_i|/a \ll 1$ for all i. Demonstrate that the equation of motion of the ith pendulum bob is
$$\ddot{\psi}_i = -\frac{g}{l}\,\psi_i + \frac{K}{m}\,(\psi_{i-1} - 2\,\psi_i + \psi_{i+1}).$$
Consider a general normal mode of the form
$$\psi_i(t) = [A\,\sin(k\,x_i) + B\,\cos(k\,x_i)]\,\cos(\omega\,t - \phi).$$
Show that the associated dispersion relation is
$$\omega^2 = \frac{g}{l} + \frac{4\,K}{m}\,\sin^2(k\,a/2).$$
Suppose that the first and last pendulums in the array are attached to immovable walls, located a horizontal distance a away, by means of light springs of



unstretched length a and spring constant K. Find the normal modes of the system. Suppose, on the other hand, that the first and last pendulums are not attached to anything on their outer sides. Find the normal modes of the system.

3. Find the system of coupled inductors and capacitors which is analogous to the system of coupled pendulums considered in the previous exercise, in the sense that the time evolution equation for the current flowing through the ith inductor has the same form as the equation of motion of the ith pendulum. Consider both types of boundary condition discussed above. Find the dispersion relation.

4. Consider a periodic waveform $y(t)$ of period τ, where $y(t+\tau) = y(t)$ for all t, which is represented as a Fourier series:

$$y(t) = C_0 + \sum_{n>1} [C_n \cos(n\,\omega\,t) + S_n \sin(n\,\omega\,t)],$$

where $\omega = 2\pi/\tau$. Demonstrate that

$$y(-t) = C_0 + \sum_{n>1} [C_n' \cos(n\,\omega\,t) + S_n' \sin(n\,\omega\,t)],$$

where $C_n' = C_n$ and $S_n' = -S_n$, and

$$y(t+T) = C_0 + \sum_{n>1} [C_n'' \cos(n\,\omega\,t) + S_n'' \sin(n\,\omega\,t)],$$

where

$$\begin{aligned} C_n'' &= C_n \cos(n\,\omega\,T) + S_n \sin(n\,\omega\,T), \\ S_n'' &= S_n \cos(n\,\omega\,T) - C_n \sin(n\,\omega\,T). \end{aligned}$$

5. Demonstrate that the periodic square-wave

$$y(t) = A \left\{ \begin{array}{ll} -1 & 0 \leq t/\tau \leq 1/2 \\ +1 & 1/2 < t/\tau \leq 1 \end{array} \right.,$$

where $y(t+\tau) = y(t)$ for all t, has the Fourier representation

$$y(t) = -\frac{4\,A}{\pi}\left[\frac{\sin(\omega\,t)}{1} + \frac{\sin(3\,\omega\,t)}{3} + \frac{\sin(5\,\omega\,t)}{5} + \cdots\right].$$

Here, $\omega = 2\pi/\tau$. Plot the reconstructed waveform, retaining the first 4, 8, 16, and 32 terms in the Fourier series.

6. Show that the periodically repeated pulse waveform

$$y(t) = A \left\{ \begin{array}{ll} 1 & |t - T/2| \leq \tau/2 \\ 0 & \text{otherwise} \end{array} \right.,$$

where $y(t+T) = y(t)$ for all t, and $\tau < T$, has the Fourier representation

$$y(t) = A\,\frac{\tau}{T} + \frac{2\,A}{\pi}\sum_{n=1,\infty}(-1)^n\,\frac{\sin(n\,\pi\,\tau/T)}{n}\,\cos(n\,2\pi\,t/T)$$

Demonstrate that if $\tau \ll T$ then the most significant terms in the above series have frequencies (in Hertz) which range from the fundamental frequency $1/T$ to a frequency of order $1/\tau$.

7 Traveling Waves

7.1 Standing Waves in a Finite Continuous Medium

We saw earlier, in Sections 5.2, 6.2, and 6.3, that a small amplitude transverse wave on a uniform string, and a small amplitude longitudinal wave in an elastic solid or an ideal gas, are all governed by the *wave equation*, which (in one dimension) takes the general form

$$\frac{\partial^2\psi}{\partial t^2} = c^2\,\frac{\partial^2\psi}{\partial x^2}, \tag{7.1}$$

where $\psi(x,t)$ represents the wave disturbance, and $c > 0$ is a constant, with the dimensions of velocity, which is a property of the particular medium that supports the wave. Up to now, we have only considered media of *finite length*: *e.g.*, media which extend from $x = 0$ to $x = l$. Generally speaking, we have encountered two distinct types of physical constraint which hold at the boundaries of such media. Firstly, if a given boundary is *fixed* then the wave displacement is constrained to be zero there: *e.g.*, if the left boundary is fixed then $\psi(0,t) = 0$. Secondly, if a given boundary is *free* then the spatial derivative of the displacement (which usually corresponds to some sort of force) is constrained to be zero there: *e.g.*, if the right boundary is free then $\partial\psi(l,t)/\partial x = 0$. It follows that a fixed boundary corresponds to a *node*—*i.e.*, a point at which the amplitude of the wave disturbance is always *zero*—whereas a free boundary corresponds to an *anti-node*—*i.e.*, a point at which the amplitude of the wave disturbance is always *locally maximal*. Consequently, the nodes and the anti-nodes of a wave, of definite wavelength, supported in a medium of finite length with stationary boundaries, which can be either fixed or free, are constrained to be *stationary*. The only simple solution of the wave equation (7.1) which has stationary nodes and anti-nodes is a *standing wave* of the general form

$$\psi(x,t) = [A\,\cos(k\,x) + B\,\sin(k\,x)]\,\cos(\omega\,t - \phi). \tag{7.2}$$

The associated nodes are located at the values of x that satisfy

$$A\,\cos(k\,x) + B\,\sin(k\,x) = 0, \tag{7.3}$$

which implies that they are indeed stationary, and also evenly spaced a distance $\lambda/2$ apart, where $\lambda = 2\pi/k$ is the wavelength. Moreover, the anti-nodes are situated halfway between the nodes. For example, suppose that both boundaries of the medium are fixed boundaries. It follows that the points $x = 0$ and $x = l$ must each correspond to a node. This is only possible if the length of the medium, l, is a *half-integer* number of wavelengths: *i.e.*, $l = n\,\lambda/2$, where n is a positive integer. We conclude that, in this case, the possible wavenumbers of standing wave solutions to the wave equation are *quantized* such that

$$k\,l = n\,\pi. \tag{7.4}$$

Moreover, the same is true if both boundaries are free boundaries. Finally, if one boundary is free, and the other fixed, then the quantization of wavenumbers takes the slightly different form

$$k\,l = (n - 1/2)\,\pi. \tag{7.5}$$

Now, those standing wave solutions that satisfy the appropriate quantization criterion are known as the *normal modes* of the system. Moreover, substitution of (7.2) into the wave equation (7.1) yields the standing wave dispersion relation

$$\omega^2 = k^2\,c^2. \tag{7.6}$$

Thus, the fact that the normal mode wavenumbers are quantized immediately implies that the associated oscillation frequencies are also quantized. Finally, since the wave equation is *linear*, the most general solution which satisfies the boundary conditions is a *linear superposition* of all of the normal modes. Such a solution has the appropriate node or anti-node at each of the boundaries, but does not necessarily have any stationary nodes or anti-nodes in the interior of the medium.

7.2 Traveling Waves in an Infinite Continuous Medium

Let us now consider solutions of the wave equation (7.1) in an *infinite* medium. Such a medium does not possess any boundaries, and so is not subject to boundary conditions. Hence, there is no particular reason why a wave of definite wavelength should have stationary nodes or anti-nodes. In other words, (7.2) may not be the only permissible type of wave solution in an infinite medium. What other kind of solution could we have? Well, suppose that

$$\psi(x,t) = A\,\cos(k\,x - \omega\,t - \phi), \tag{7.7}$$

where $A > 0$, $k > 0$, $\omega > 0$, and ϕ are constants. We would interpret this solution as a wave of amplitude A, wavenumber k, wavelength $\lambda = 2\pi/k$, angular frequency ω, frequency (in Hertz) $f = \omega/2\pi$, period $T = 1/f$, and phase angle ϕ. Note, in particular, that $\psi(x+\lambda, t) = \psi(x, t)$ and $\psi(x, t+T) = \psi(x, t)$ for all x and t: *i.e.*, the wave is periodic in space with period λ, and periodic in time with period T. Now, a *wave maximum* corresponds to a point at which $\cos(k\,x - \omega\,t - \phi) = 1$. It follows, from the well known properties of the cosine function, that the various wave maxima are located at

$$k\,x - \omega\,t - \phi = n\,2\pi, \tag{7.8}$$

where n is an integer. Thus, differentiating the above expression with respect to t, and rearranging, the equation of motion of a particular maximum becomes

$$\frac{dx}{dt} = \frac{\omega}{k}. \tag{7.9}$$

We conclude that the wave maximum in question *propagates* along the x-axis at the velocity

$$v_p = \frac{\omega}{k}. \tag{7.10}$$

It is easily demonstrated that the other wave maxima, as well as the wave minima and the wave zeros, also propagate along the x-axis at the same velocity. In fact, the whole wave pattern propagates in the positive x-direction *without changing shape*. The characteristic propagation velocity v_p is known as the *phase velocity* of the wave, since it is the velocity with which points of *constant phase* in the wave disturbance (*i.e.*, points which satisfy $k\,x - \omega\,t - \phi =$ constant) move. For obvious reasons, the type of wave solution given in (7.7) is called a *traveling wave*.

Substitution of (7.7) into the wave equation (7.1) yields the familiar dispersion relation

$$\omega^2 = k^2\,c^2. \tag{7.11}$$

We immediately conclude that the traveling wave solution (7.7) satisfies the wave equation provided

$$v_p = \frac{\omega}{k} = c: \tag{7.12}$$

i.e., provided that the phase velocity of the wave takes the fixed value c. In other words, the constant c^2, which appears in the wave equation (7.1), can be interpreted as the square of the velocity with which traveling waves propagate through the medium in question. It follows, from the discussion in Sections 5.2, 6.2, and 6.3, that transverse waves propagate along strings of tension T and mass per unit length ρ at the phase velocity $\sqrt{T/\rho}$, that longitudinal sound waves propagate through elastic media of Young's modulus Y and mass density ρ at the phase velocity $\sqrt{Y/\rho}$, and that sound waves

Material	$Y\,(\mathrm{N\,m^{-2}})$	$\rho\,(\mathrm{kg\,m^{-3}})$	$\sqrt{Y/\rho}\,(\mathrm{m\,s^{-1}})$	$v\,(\mathrm{m\,s^{-1}})$
Aluminium	6.0×10^{10}	2.7×10^{3}	4700	5100
Granite	5.0×10^{10}	2.7×10^{3}	4300	~ 5000
Lead	$\sim 1.6 \times 10^{10}$	11.4×10^{3}	1190	1320
Nickel	21.4×10^{10}	8.9×10^{3}	4900	4970
Pyrex	6.1×10^{10}	2.25×10^{3}	5200	5500
Silver	7.5×10^{10}	10.4×10^{3}	2680	2680

Table 7.1: *Calculated versus measured sound velocities in various solid materials. [From* Vibrations and Waves*, A.P. French (W.W. Norton & Co., New York NY, 1971).]*

propagate through ideal gases of pressure p, mass density ρ, and ratio of specific heats γ, at the phase velocity $\sqrt{\gamma\,p/\rho}$.

Table 7.1 shows some data on the calculated and measured speeds of sound in various solid materials. It can be seen that the agreement between the two is fairly good. Actually, the formula $v = \sqrt{Y/\rho}$ is only valid if the material in question is free to (very slightly) expand and contract sideways as a wave of compression or decompression passes by. However, bulk material is not free to do this, and so its resistance to deformation is effectively increased. This, typically, has the effect of raising the sound speed by about 15%.

An *ideal gas* of mass m and molecular weight M satisfies the *ideal gas equation of state*,

$$p\,V = \frac{m}{M}\,R\,T, \tag{7.13}$$

where p is the pressure, V the volume, R the gas constant, and T the absolute temperature. Since the ratio m/V is equal to the density, ρ, the expression for the sound speed, $v = \sqrt{\gamma\,p/\rho}$, yields

$$v = \left(\frac{\gamma\,R\,T}{M}\right)^{1/2}. \tag{7.14}$$

We, thus, conclude that the speed of sound in an ideal gas is independent of the pressure or the density, proportional to the square root of the absolute temperature, and inversely proportional to the square root of the molecular mass. All of these predictions are borne out in practice.

A comparison of Equations (7.6) and (7.11) reveals that standing waves and traveling waves in a given medium satisfy the *same* dispersion relation.

However, since traveling waves in infinite media are not subject to boundary conditions, it follows that there is *no restriction* on the possible wavenumbers, or wavelengths, of such waves. Hence, *any* traveling wave solution whose wavenumber, k, and angular frequency, ω, are related according to

$$\omega = k\,c \tag{7.15}$$

[which is just the square root of the dispersion relation (7.11)] is a valid solution of the wave equation. Another way of putting this is that any traveling wave solution whose wavelength, $\lambda = 2\pi/k$, and frequency, $f = \omega/2\pi$, are related according to

$$c = f\,\lambda \tag{7.16}$$

is a valid solution of the wave equation. We, thus, conclude that high frequency traveling waves propagating through a given medium possess short wavelengths, and *vice versa*.

Consider the alternative wave solution

$$\psi(x,t) = A\,\cos(k\,x + \omega\,t - \phi), \tag{7.17}$$

where $A > 0$, $k > 0$, $\omega > 0$, and ϕ are constants. As before, we would interpret this solution as a wave of amplitude A, wavenumber k, angular frequency ω, and phase angle ϕ. However, the wave maxima are now located at

$$k\,x + \omega\,t - \phi = n\,2\pi, \tag{7.18}$$

where n is an integer, and thus have equations of motion of the form

$$\frac{dx}{dt} = -\frac{\omega}{k}. \tag{7.19}$$

Clearly, (7.17) represents a traveling wave that propagates in the *minus* x-direction at the phase velocity $v_p = \omega/k$. Moreover, substitution of (7.17) into the wave equation (7.1) again yields the dispersion relation (7.15), which implies that $v_p = c$. It follows that traveling wave solutions to the wave equation (7.1) can propagate in either the positive or the negative x-direction, as long as they always move at the fixed speed c.

7.3 Wave Interference

But, what is the relationship between traveling wave and standing wave solutions to the wave equation (7.1) in an infinite medium? To help answer this question, let us form a superposition of two traveling wave solutions of equal amplitude A, and zero phase angle ϕ, which have the same wavenumber k, but are moving in *opposite directions*. In other words,

$$\psi(x,t) = A\,\cos(k\,x-\omega\,t) + A\,\cos(k\,x+\omega\,t). \tag{7.20}$$

Since the wave equation (7.1) is linear, it follows that the above superposition is a solution provided the two component waves are also solutions: *i.e.*, provided that $\omega = k\,c$, which we shall assume to be the case. However, making use of the trigonometric identity $\cos a + \cos b \equiv 2\,\cos[(a+b)/2]\,\cos[(a-b)/2]$, the above expression can also be written

$$\psi(x,t) = 2\,A\,\cos(k\,x)\,\cos(\omega\,t), \tag{7.21}$$

which is clearly a standing wave [*cf.*, (7.2)]. Evidently, a standing wave is a linear superposition of two, otherwise identical, traveling waves which propagate in opposite directions. The two waves completely cancel one another out at the nodes, which are situated at $k\,x = (n-1/2)\,\pi$, where n is an integer. This process is known as *total destructive interference*. On the other hand, the waves reinforce one another at the anti-nodes, which are situated at $k\,x = n\,\pi$, generating a wave whose amplitude is twice that of the component waves. This process is known as *constructive interference*.

As a more general example of wave interference, consider a superposition of two traveling waves of unequal amplitudes which again have the same wavenumber and zero phase angle, and are moving in opposite directions: *i.e.*,

$$\psi(x,t) = A_1\,\cos(k\,x-\omega\,t) + A_2\,\cos(k\,x+\omega\,t), \tag{7.22}$$

where $A_1, A_2 > 0$. In this case, the trigonometric identities $\cos(a-b) \equiv \cos a\,\cos b + \sin a\,\sin b$ and $\cos(a+b) \equiv \cos a\,\cos b - \sin a\,\sin b$ yield

$$\psi(x,t) = (A_1+A_2)\,\cos(k\,x)\,\cos(\omega\,t) + (A_1-A_2)\,\sin(k\,x)\,\sin(\omega\,t). \tag{7.23}$$

Thus, the two waves interfere destructively at $k\,x = (n-1/2)\,\pi$ [*i.e.*, at points where $\cos(k\,x) = 0$ and $|\sin(k\,x)| = 1$] to produce a minimum wave amplitude $|A_1 - A_2|$, and interfere constructively at $k\,x = n\,\pi$ [*i.e.*, at points where $|\cos(k\,x)| = 1$ and $\sin(k\,x) = 0$] to produce a maximum wave amplitude A_1+A_2. Note, however, that the destructive interference is incomplete unless $A_1 = A_2$. Incidentally, it is a general result that when two waves of amplitude $A_1 > 0$ and $A_2 > 0$ interfere then the maximum and minimum possible values of the resulting wave amplitude are $A_1 + A_2$ and $|A_1 - A_2|$, respectively.

7.4 Energy Conservation

Consider a small amplitude transverse wave propagating along a uniform string of infinite length, tension T, and mass per unit length ρ. See Section 5.2. Let x measure distance along the string, and let $y(x,t)$ be the transverse wave displacement. Of course, $y(x,t)$ satisfies the wave equation

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2}, \tag{7.24}$$

where $v = \sqrt{T/\rho}$ is the phase velocity of traveling waves on the string.

Consider a section of the string lying between $x = x_1$ and $x = x_2$. The *kinetic energy* of this section is

$$K = \int_{x_1}^{x_2} \frac{1}{2}\,\rho \left(\frac{\partial y}{\partial t}\right)^2 dx, \tag{7.25}$$

since $\partial y/\partial t$ is the string's transverse velocity. The potential energy is simply the work done in stretching the section, which is $T\,\Delta s$, where Δs is the difference between the section's stretched and unstretched lengths. Here, it is assumed that the tension remains approximately constant when the section is slowly stretched. Now, an element of length of the string is

$$ds = (dx^2 + dy^2)^{1/2} = \left[1 + \left(\frac{dy}{dx}\right)^2\right]^{1/2} dx. \tag{7.26}$$

Hence,

$$\Delta s = \int_{x_1}^{x_2} \left\{\left[1 + \left(\frac{dy}{dx}\right)^2\right]^{1/2} - 1\right\} dx \simeq \int_{x_1}^{x_2} \frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2 dx, \tag{7.27}$$

since it is assumed that $|\partial y/\partial x| \ll 1$: *i.e.*, the transverse displacement is sufficiently small that the string remains almost parallel to the x-axis. Thus, the potential energy of the section is $U = T\,\Delta s$, or

$$U = \int_{x_1}^{x_2} \frac{1}{2}\,T \left(\frac{\partial y}{\partial x}\right)^2 dx. \tag{7.28}$$

It follows that the total energy of the section is

$$E = \int_{x_1}^{x_2} \frac{1}{2}\left[\rho\left(\frac{\partial y}{\partial t}\right)^2 + T\left(\frac{\partial y}{\partial x}\right)^2\right] dx. \tag{7.29}$$

Multiplying the wave equation (7.24) by $\rho\,(\partial y/\partial t)$, we obtain

$$\rho\,\frac{\partial y}{\partial t}\,\frac{\partial^2 y}{\partial t^2} = T\,\frac{\partial y}{\partial t}\,\frac{\partial^2 y}{\partial x^2}, \tag{7.30}$$

since $v^2 = T/\rho$. This expression yields

$$\rho\,\frac{\partial y}{\partial t}\,\frac{\partial^2 y}{\partial t^2} + T\,\frac{\partial y}{\partial x}\,\frac{\partial^2 y}{\partial t\,\partial x} = T\,\frac{\partial y}{\partial t}\,\frac{\partial^2 y}{\partial x^2} + T\,\frac{\partial y}{\partial x}\,\frac{\partial^2 y}{\partial t\,\partial x}, \tag{7.31}$$

which can be written in the form

$$\frac{1}{2}\,\frac{\partial}{\partial t}\left[\rho\left(\frac{\partial y}{\partial t}\right)^2 + T\left(\frac{\partial y}{\partial x}\right)^2\right] = \frac{\partial}{\partial x}\left(T\,\frac{\partial y}{\partial t}\,\frac{\partial y}{\partial x}\right), \tag{7.32}$$

or

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial \mathcal{I}}{\partial x} = 0, \tag{7.33}$$

where

$$\mathcal{E}(x,t) = \frac{1}{2}\left[\rho\left(\frac{\partial y}{\partial t}\right)^2 + T\left(\frac{\partial y}{\partial x}\right)^2\right] \tag{7.34}$$

is the *energy per unit length* of the string, and

$$\mathcal{I}(x,t) = -T\,\frac{\partial y}{\partial t}\,\frac{\partial y}{\partial x}. \tag{7.35}$$

Finally, integrating (7.33) in x from x_1 to x_2, we obtain

$$\frac{d}{dt}\int_{x_1}^{x_2} \mathcal{E}\,dx + \mathcal{I}(x_2,t) - \mathcal{I}(x_1,t) = 0, \tag{7.36}$$

or

$$\frac{dE}{dt} = \mathcal{I}(x_1,t) - \mathcal{I}(x_2,t). \tag{7.37}$$

Here, $E(t)$ is the energy stored in the section of the string lying between $x = x_1$ and $x = x_2$ [see Equation (7.29)]. Now, if we interpret $\mathcal{I}(x,t)$ as the instantaneous *energy flux* (*i.e.*, rate of energy flow) in the positive-x direction, at position x and time t, then the above equation can be recognized as a simple declaration of *energy conservation*. Basically, the equation states that the rate of increase in the energy stored in the section of the string lying between $x = x_1$ and $x = x_2$, which is dE/dt, is equal to the difference between the rate at which energy flows into the left end of the section, which is $\mathcal{I}(x_1,t)$, and the rate at which it flows out of the right end, which is $\mathcal{I}(x_2,t)$.

Note that the string *must* conserve energy, since it lacks any mechanism for energy dissipation. The same is true of the other wave media discussed in this section.

Consider a wave propagating in the positive x-direction of the form

$$y(x,t) = A\,\cos(k\,x - \omega\,t - \phi). \tag{7.38}$$

According to Equation (7.35), the energy flux associated with this wave is

$$\mathcal{I}(x,t) = T\,k\,\omega\,A^2\,\sin^2(k\,x - \omega\,t - \phi). \tag{7.39}$$

Thus, the mean energy flux is written

$$\langle \mathcal{I}\rangle = \frac{1}{2}\,\omega^2\,Z\,A^2, \tag{7.40}$$

where $\langle A\rangle(x) \equiv (\omega/2\pi)\int_t^{t+2\pi/\omega} A(x,t')\,dt'$ represents an average over a period of the wave oscillation. Here, use has been made of $\omega/k = \sqrt{T/\rho}$, and the easily demonstrated result that $\langle\sin^2(\omega\,t + \theta)\rangle = 1/2$ for all θ. Moreover, the quantity

$$Z = \sqrt{\rho\,T} \tag{7.41}$$

is known as the characteristic *impedance* of the string. The units of Z are force over velocity. Thus, the string impedance measures the typical tension required to produce a unit transverse velocity. Finally, according to Equation (7.40), a traveling wave propagating in the positive x-direction is associated with a positive energy flux: *i.e.*, the wave *transports energy* in the positive x-direction.

Consider a wave propagating in the negative x-direction of the general form

$$y(x,t) = A\,\cos(k\,x + \omega\,t - \phi). \tag{7.42}$$

It is easily demonstrated, from (7.35), that the mean energy flux associated with this wave is

$$\langle \mathcal{I}\rangle = -\frac{1}{2}\,\omega^2\,Z\,A^2. \tag{7.43}$$

The fact that the energy flux is negative means that the wave transports energy in the negative x-direction.

Suppose that we have a superposition of a wave of amplitude A_+ propagating in the positive x-direction, and a wave of amplitude A_- propagating in the negative x-direction, so that

$$y(x,t) = A_+\,\cos(k\,x - \omega\,t - \phi_+) + A_-\,\cos(k\,x + \omega\,t - \phi_-). \tag{7.44}$$

According to (7.35), the instantaneous energy flux is written

$$\begin{aligned}\mathcal{I}(x,t) &= \omega^2 Z\,[A_+ \sin(k\,x-\omega\,t-\phi_+) + A_- \sin(k\,x+\omega\,t-\phi_-)]\\ &\quad [A_+ \sin(k\,x-\omega\,t-\phi_+) - A_- \sin(k\,x+\omega\,t-\phi_-)]\\ &= \omega^2 Z\left[A_+^2 \sin^2(k\,x-\omega\,t-\phi_+) - A_-^2 \sin^2(k\,x+\omega\,t-\phi_-)\right].\end{aligned} \tag{7.45}$$

Hence, the mean energy flux,

$$\langle\mathcal{I}\rangle = \frac{1}{2}\,\omega^2 Z\,A_+^2 - \frac{1}{2}\,\omega^2 Z\,A_-^2, \tag{7.46}$$

is simply the difference between the mean fluxes associated with the waves traveling to the right (*i.e.*, in the positive x-direction) and to the left, calculated separately. Recall, from the previous section, that a standing wave is a superposition of two traveling waves of *equal amplitude*, and frequency, propagating in opposite directions. It immediately follows, from the above expression, that a standing wave has *zero* associated net energy flux. In other words, a standing wave does not give rise to net energy transport.

Now, we saw earlier, in Section 6.2, that a small amplitude longitudinal wave in an elastic solid satisfies the wave equation,

$$\frac{\partial^2\psi}{\partial t^2} = c^2\,\frac{\partial^2\psi}{\partial x^2}, \tag{7.47}$$

where $\psi(x,t)$ is the longitudinal wave displacement, $c = \sqrt{Y/\rho}$ the phase velocity of traveling waves in the solid, Y the Young's modulus, and ρ the mass density. Using similar analysis to that employed above, we can derive an energy conservation equation of the form (7.33) from the above wave equation, where

$$\mathcal{E} = \frac{1}{2}\left[\rho\left(\frac{\partial\psi}{\partial t}\right)^2 + Y\left(\frac{\partial\psi}{\partial x}\right)^2\right] \tag{7.48}$$

is the total wave energy per unit volume, and

$$\mathcal{I} = -Y\,\frac{\partial\psi}{\partial t}\,\frac{\partial\psi}{\partial x} \tag{7.49}$$

the wave energy flux (*i.e.*, rate of energy flow per unit area) in the positive x-direction. For a traveling wave of the form $\psi(x,t) = A\,\cos(k\,x-\omega\,t-\phi)$, the above expression yields

$$\langle\mathcal{I}\rangle = \frac{1}{2}\,\omega^2 Z\,A^2, \tag{7.50}$$

where

$$Z = \sqrt{\rho\, Y} \tag{7.51}$$

is the impedance of the solid. The units of Z are pressure over velocity, so, in this case, the impedance measures the typical pressure in the solid required to produce a unit longitudinal velocity. Analogous arguments to the above reveal that the impedance of an ideal gas of density ρ, pressure p, and ratio of specific heats γ, is (see Section 6.3)

$$Z = \sqrt{\rho\, \gamma\, p}. \tag{7.52}$$

7.5 Transmission Lines

A *transmission line* is typically used to carry high frequency electromagnetic signals over large distances: *i.e.*, distances sufficiently large that the phase of the signal varies significantly along the line (which implies that the line is much longer than the wavelength of the signal). A common example of a transmission line is an ethernet cable. In its simplest form, a transmission line consists of two parallel conductors which carry equal and opposite electrical currents $I(x,t)$, where x measures distance along the line. Let $V(x,t)$ be the instantaneous voltage difference between the two conductors at position x. Consider a small section of the line lying between x and $x+\delta x$. If $Q(t)$ is the electric charge on one of the conducting sections, and $-Q(t)$ the charge on the other, then charge conservation implies that $dQ/dt = I(x,t) - I(x+\delta x,t)$. However, according to standard electrical circuit theory, $Q(t) = \mathcal{C}\,\delta x\, V(x,t)$, where $\mathcal{C}$ is the *capacitance per unit length* of the line. Standard circuit theory also yields $V(x+\delta x,t) - V(x,t) = -\mathcal{L}\,\delta x\,\partial I(x,t)/\partial t$, where $\mathcal{L}$ is the *inductance per unit length* of the line. Taking the limit $\delta x \to 0$, we obtain the so-called *Telegrapher's equations*,

$$\frac{\partial V}{\partial t} = -\frac{1}{\mathcal{C}}\,\frac{\partial I}{\partial x}, \tag{7.53}$$

$$\frac{\partial I}{\partial t} = -\frac{1}{\mathcal{L}}\,\frac{\partial V}{\partial x}. \tag{7.54}$$

These two equations can be combined to give

$$\frac{\partial^2 V}{\partial t^2} = \frac{1}{\mathcal{L}\,\mathcal{C}}\,\frac{\partial^2 V}{\partial x^2}, \tag{7.55}$$

together with an analogous equation for I. In other words, $V(x,t)$ and $I(x,t)$ both obey a wave equation of the form (7.24) in which the associated phase velocity is $v = 1/\sqrt{\mathcal{L}\,\mathcal{C}}$. Multiplying (7.53) by $\mathcal{C}\,V$, (7.54) by $\mathcal{L}\,I$, and then adding the two resulting expressions, we obtain the energy conservation equation

$$\frac{\partial\mathcal{E}}{\partial t}+\frac{\partial\mathcal{I}}{\partial x}=0, \tag{7.56}$$

where

$$\mathcal{E}=\frac{1}{2}\,\mathcal{L}\,I^2+\frac{1}{2}\,\mathcal{C}\,V^2 \tag{7.57}$$

is the electromagnetic *energy per unit length* of the line, and

$$\mathcal{I}=I\,V \tag{7.58}$$

is the electromagnetic *energy flux* along the line (*i.e.*, the energy per unit time which passes a given point) in the positive x-direction. Consider a signal propagating along the line, in the positive x-direction, whose associated current takes the form

$$I(x,t)=I_0\,\cos(k\,x-\omega\,t-\phi). \tag{7.59}$$

It is easily demonstrated, from (7.53), that the corresponding voltage is

$$V(x,t)=V_0\,\cos(k\,x-\omega\,t-\phi), \tag{7.60}$$

where

$$V_0=I_0\,Z. \tag{7.61}$$

Here,

$$Z=\sqrt{\frac{\mathcal{L}}{\mathcal{C}}} \tag{7.62}$$

is the characteristic *impedance* of the line, and is measured in Ohms. It follows that the mean energy flux associated with the signal is written

$$\langle\mathcal{I}\rangle=\langle I\,V\rangle=\frac{1}{2}\,I_0\,V_0=\frac{1}{2}\,Z\,I_0^{\,2}=\frac{1}{2}\,\frac{V_0^{\,2}}{Z}. \tag{7.63}$$

Likewise, for a signal propagating along the line in the negative x-direction,

$$\begin{aligned} I(x,t) &= I_0\,\cos(k\,x+\omega\,t-\phi), && (7.64)\\ V(x,t) &= -V_0\,\cos(k\,x+\omega\,t-\phi), && (7.65) \end{aligned}$$

and the mean energy flux is

$$\langle\mathcal{I}\rangle=-\frac{1}{2}\,I_0\,V_0=-\frac{1}{2}\,Z\,I_0^{\,2}=-\frac{1}{2}\,\frac{V_0^{\,2}}{Z}. \tag{7.66}$$

As a specific example, consider a transmission line consisting of two uniform parallel conducting strips of width w and perpendicular distance apart d, where $d \ll w$. It is easily demonstrated, using elementary electrostatic theory, that the capacitance per unit length of the line is

$$\mathcal{C} = \epsilon_0 \, \frac{w}{d}, \tag{7.67}$$

where $\epsilon_0 = 8.8542 \times 10^{-12}\,\mathrm{C^2\,N^{-1}\,m^{-2}}$ is the *electric permittivity of free space*. Likewise, according to elementary magnetostatic theory, the line's inductance per unit length takes the form

$$\mathcal{L} = \mu_0 \, \frac{d}{w}, \tag{7.68}$$

where $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{N\,A^{-2}}$ is the *magnetic permeability of free space*. Thus, the phase velocity of a signal propagating down the line is

$$v = \frac{1}{\sqrt{\mathcal{L}\,\mathcal{C}}} = \frac{1}{\sqrt{\epsilon_0\,\mu_0}} = 2.998 \times 10^8\,\mathrm{m\,s^{-1}}, \tag{7.69}$$

which, of course, is the *velocity of light* in vacuum [see Equation (7.113)]. Furthermore, the impedance of the line is

$$Z = \sqrt{\frac{\mathcal{L}}{\mathcal{C}}} = \frac{d}{w}\, Z_0, \tag{7.70}$$

where the quantity

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 376.73\,\Omega \tag{7.71}$$

is known as the *impedance of free space*.

7.6 Reflection and Transmission at Boundaries

Consider two uniform semi-infinite strings which run along the x-axis, and are tied together at $x = 0$. Let the first string be of density per unit length ρ_1, and occupy the region $x < 0$, and let the second string be of density per unit length ρ_2, and occupy the region $x > 0$. Now, the tensions in the two strings must be *equal*, otherwise the string interface would not be in force balance in the x-direction. So, let T be the common tension. Suppose that a transverse wave of angular frequency ω is launched from a wave source at $x = -\infty$, and propagates towards the interface. Assuming that $\rho_1 \neq \rho_2$, we would expect the wave incident on the interface to be partially *reflected*, and partially *transmitted*. Of course, the frequencies of the incident, reflected, and transmitted waves are all the *same*, since this property of the waves is ultimately determined by the oscillation frequency of the wave source. Hence, in the region $x < 0$, the wave displacement takes the form

$$y(x,t) = A_i \cos(k_1 x - \omega t - \phi_i) + A_r \cos(k_1 x + \omega t - \phi_r). \tag{7.72}$$

In other words, the displacement is a linear superposition of an *incident wave* and a *reflected wave*. The incident wave propagates in the positive x-direction, and is of amplitude A_i, wavenumber $k_1 = \omega/v_1$, and phase angle ϕ_i. The reflected wave propagates in the negative x-direction, and is of amplitude A_r, wavenumber $k_1 = \omega/v_1$, and phase angle ϕ_r. Here, $v_1 = \sqrt{T/\rho_1}$ is the phase velocity of traveling waves on the first string. In the region $x > 0$, the wave displacement takes the form

$$y(x,t) = A_t \cos(k_2 x - \omega t - \phi_t). \tag{7.73}$$

In other words, the displacement is solely due to a *transmitted wave* which propagates in the positive x-direction, and is of amplitude A_t, wavenumber $k_2 = \omega/v_2$, and phase angle ϕ_t. Here, $v_2 = \sqrt{T/\rho_2}$ is the phase velocity of traveling waves on the second string.

Let us now consider the matching conditions at the interface between the two strings: *i.e.*, at $x = 0$. Firstly, since the two strings are tied together at $x = 0$, their transverse displacements at this point must be equal to one another. In other words,

$$y(0_-, t) = y(0_+, t), \tag{7.74}$$

or

$$A_i \cos(\omega t + \phi_i) + A_r \cos(\omega t - \phi_r) = A_t \cos(\omega t + \phi_t). \tag{7.75}$$

The only way in which the above equation can be satisfied for all values of t is if $\phi_i = -\phi_r = \phi_t$. This being the case, the common $\cos(\omega t + \phi_i)$ factor cancels out, and we are left with

$$A_i + A_r = A_t. \tag{7.76}$$

Secondly, since the two strings lack an energy dissipation mechanism, the energy flux into the interface must match that out of the interface. In other words,

$$\frac{1}{2}\,\omega^2 Z_1\,(A_i^2 - A_r^2) = \frac{1}{2}\,\omega^2 Z_2\, A_t^2, \tag{7.77}$$

where $Z_1 = \sqrt{\rho_1 T}$ and $Z_2 = \sqrt{\rho_2 T}$ are the impedances of the first and second strings, respectively. The above expression reduces to

$$Z_1 (A_i + A_r)(A_i - A_r) = Z_2 A_t^2, \tag{7.78}$$

which, when combined with Equation (7.76), yields

$$Z_1 (A_i - A_r) = Z_2 A_t. \tag{7.79}$$

Equations (7.76) and (7.79) can be solved to give

$$A_r = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right) A_i, \tag{7.80}$$

$$A_t = \left(\frac{2 Z_1}{Z_1 + Z_2}\right) A_i. \tag{7.81}$$

The *coefficient of reflection*, R, is defined as the ratio of the reflected to the incident energy flux: *i.e.*,

$$R = \left(\frac{A_r}{A_i}\right)^2 = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2. \tag{7.82}$$

The *coefficient of transmission*, T, is defined as the ratio of the transmitted to the incident energy flux: *i.e.*,

$$T = \frac{Z_2}{Z_1}\left(\frac{A_t}{A_i}\right)^2 = \frac{4 Z_1 Z_2}{(Z_1 + Z_2)^2}. \tag{7.83}$$

Note that

$$R + T = 1: \tag{7.84}$$

i.e., any incident wave energy which is not reflected is transmitted.

Suppose that the density per unit length of the second string, ρ_2, tends to infinity, so that $Z_2 = \sqrt{\rho_2 T} \to \infty$. It follows from (7.80) and (7.81) that $A_r = -A_i$ and $A_t = 0$. Likewise, (7.82) and (7.83) yield $R = 1$ and $T = 0$. Hence, the interface between the two strings is *stationary* (since it oscillates with amplitude A_t), and there is no transmitted energy. In other words, the second string acts exactly like a *fixed boundary*. It follows that when a transverse wave on a string is incident on a fixed boundary then it is perfectly reflected with a phase shift of π: *i.e.*, $A_r = -A_i$. Thus, the resultant wave displacement on the string becomes

$$\begin{aligned} y(x,t) &= A_i \cos(k_1 x - \omega t - \phi_i) - A_i \cos(k_1 x + \omega t + \phi_i) \\ &= 2 A_i \sin(k_1 x) \sin(\omega t + \phi_i), \end{aligned} \tag{7.85}$$

where use has been made of the trigonometric identity $\cos a - \cos b \equiv 2 \sin[(a+b)/2] \sin[(b-a)/2]$. We conclude that the incident and reflected waves interfere in such a manner as to produce a *standing wave* with a *node* at the fixed boundary.

Suppose that the density per unit length of the second string, ρ_2, tends to zero, so that $Z_2 = \sqrt{\rho_2 T} \rightarrow 0$. It follows from (7.80) and (7.81) that $A_r = A_i$ and $A_t = 2 A_i$. Likewise, (7.82) and (7.83) yield $R = 1$ and $T = 0$. Hence, the interface between the two strings oscillates at twice the amplitude of the incident wave (*i.e.*, the interface is a point of maximal amplitude oscillation), and there is no transmitted energy. In other words, the second string acts exactly like a *free boundary*. It follows that when a transverse wave on a string is incident on a free boundary then it is perfectly reflected with no phase shift: *i.e.*, $A_r = A_i$. Thus, the resultant wave displacement on the string becomes

$$\begin{aligned} y(x,t) &= A_i \cos(k_1 x - \omega t - \phi_i) + A_i \cos(k_1 x + \omega t + \phi_i) \\ &= 2 A_i \cos(k_1 x) \cos(\omega t + \phi_i), \end{aligned} \tag{7.86}$$

where use has been made of the trigonometric identity $\cos a + \cos b \equiv 2 \cos[(a+b)/2] \cos[(a-b)/2]$. We conclude that the incident and reflected waves interfere in such a manner as to produce a *standing wave* with an *anti-node* at the free boundary.

Suppose that two strings of mass per unit length ρ_1 and ρ_2 are separated by a short section of string of mass per unit length ρ_3. Let all three strings have the common tension T. Suppose that the first and second strings occupy the regions $x < 0$ and $x > a$, respectively. Thus, the middle string occupies the region $0 \leq x \leq a$. Moreover, the interface between the first and middle strings is at $x = 0$, and the interface between the middle and second strings is at $x = a$. Suppose that a wave of angular frequency ω is launched from a wave source at $x = -\infty$, and propagates towards the two interfaces. We would expect this wave to be partially reflected and partially transmitted at the first interface ($x = 0$), and the resulting transmitted wave to then be partially reflected and partially transmitted at the second interface ($x = a$). Thus, we can write the wave displacement in the region $x < 0$ as

$$y(x,t) = A_i \cos(k_1 x - \omega t) + A_r \cos(k_1 x + \omega t), \tag{7.87}$$

where A_i is the amplitude of the incident wave, A_r is the amplitude of the reflected wave, and $k_1 = \omega/\sqrt{T/\rho_1}$. Here, the phase angles of the two waves have been chosen so as to facilitate the matching process at $x = 0$.

The wave displacement in the region $x > a$ takes the form

$$y(x,t) = A_t \cos(k_2 x - \omega t - \phi_t), \tag{7.88}$$

where A_t is the amplitude of the final transmitted wave, and $k_2 = \omega/\sqrt{T/\rho_2}$. Finally, the wave displacement in the region $0 \leq x \leq a$ is written

$$y(x,t) = A_+ \cos(k_3 x - \omega t) + A_- \cos(k_3 x + \omega t), \tag{7.89}$$

where A_+ and A_- are the amplitudes of the right and left moving waves on the middle string, respectively, and $k_3 = \omega/\sqrt{T/\rho_3}$. Continuity of the transverse displacement at $x = 0$ yields

$$A_i + A_r = A_+ + A_-, \tag{7.90}$$

where a common factor $\cos(\omega t)$ has cancelled out. Continuity of the energy flux at $x = 0$ gives

$$Z_1 (A_i^2 - A_r^2) = Z_3 (A_+^2 - A_-^2), \tag{7.91}$$

so the previous two expressions can be combined to produce

$$Z_1 (A_i - A_r) = Z_3 (A_+ - A_-). \tag{7.92}$$

Continuity of the transverse displacement at $y = a$ yields

$$A_+ \cos(k_3 a - \omega t) + A_- \cos(k_3 a + \omega t) = A_t \cos(k_2 a - \omega t - \phi_t). \tag{7.93}$$

Suppose that the length of the middle string is *one quarter of a wavelength*: *i.e.*, $k_3 a = \pi/2$. Furthermore, let $\phi_t = k_2 a + \pi/2$. It follows that $\cos(k_3 a - \omega t) = \sin(\omega t)$, $\cos(k_3 a + \omega t) = -\sin(\omega t)$, and $\cos(k_2 a - \omega t - \phi_t) = \sin(\omega t)$. Thus, canceling out a common factor $\sin(\omega t)$, the above expression yields

$$A_+ - A_- = A_t. \tag{7.94}$$

Continuity of the energy flux at $x = a$ gives

$$Z_3 (A_+^2 - A_-^2) = Z_2 A_t^2. \tag{7.95}$$

so the previous two equations can be combined to generate

$$Z_3 (A_+ + A_-) = Z_2 A_t. \tag{7.96}$$

Equations (7.90) and (7.96) yield

$$A_i + A_r = \frac{Z_2}{Z_3} A_t, \tag{7.97}$$

whereas Equations (7.92) and (7.94) give

$$A_i - A_r = \frac{Z_3}{Z_1}\,A_t, \tag{7.98}$$

so, combining the previous two expression, we obtain

$$A_r = \left(\frac{Z_1\,Z_2 - Z_3^2}{Z_1\,Z_2 + Z_3^2}\right) A_i, \tag{7.99}$$

$$A_t = \left(\frac{2\,Z_1\,Z_3}{Z_1\,Z_2 + Z_3^2}\right) A_i. \tag{7.100}$$

Finally, the overall coefficient of reflection is

$$R = \left(\frac{A_r}{A_i}\right)^2 = \left(\frac{Z_1\,Z_2 - Z_3^2}{Z_1\,Z_2 + Z_3^2}\right)^2, \tag{7.101}$$

whereas the overall coefficient of transmission becomes

$$T = \frac{Z_2}{Z_1}\left(\frac{A_t}{A_i}\right)^2 = \frac{4\,Z_1\,Z_2\,Z_3^2}{(Z_1\,Z_2 + Z_3^2)^2} = 1 - R. \tag{7.102}$$

Now, suppose that the impedance of the middle string is the *geometric mean* of the impedances of the two outer strings: *i.e.*, $Z_3 = \sqrt{Z_1\,Z_2}$. In this case, it is clear, from the above two equations, that $R = 0$ and $T = 1$. In other words, there is *no reflection* of the incident wave, and all of the incident energy ends up being transmitted across the middle string from the leftmost to the rightmost string. Thus, if we wish to transmit transverse wave energy from a string of impedance Z_1 to a string of impedance Z_2 (where $Z_2 \neq Z_1$) in the most efficient manner possible—*i.e*, with no reflection of the incident energy flux—then we can do this by connecting the two strings via a short section of string whose length is one quarter of a wavelength, and whose impedance is $\sqrt{Z_1\,Z_2}$. This procedure is known as *impedance matching*.

It should be reasonably clear that the above analysis of the reflection and transmission of transverse waves at a boundary between two strings is also applicable to the reflection and transmission of other types of wave incident on a boundary between two media of differing impedances. For example, consider a *transmission line*, such as a co-axial cable. Suppose that the line occupies the region $x < 0$, and is terminated (at $x = 0$) by a load resistor of resistance R_L. Such a resistor might represent a radio antenna (which acts just like a resistor in an electrical circuit, except that the dissipated energy is radiated, rather than being converted into heat energy). Suppose that a signal of angular frequency ω is sent down the line from a wave source at $x = -\infty$. The current and voltage on the line can be written

$$I(x, t) = I_i \cos(k\,x - \omega\,t) + I_r \cos(k\,x + \omega\,t), \tag{7.103}$$

$$V(x, t) = I_i\,Z \cos(k\,x - \omega\,t) - Z\,I_r \cos(k\,x + \omega\,t), \tag{7.104}$$

where I_i is the amplitude of the incident signal, I_r the amplitude of the signal reflected by the load, Z the characteristic impedance of the line, and $k = \omega/v$. Here, v is the characteristic phase velocity with which signals propagate down the line. See Section 7.5. Now, the resistor obeys Ohm's law, which yields

$$V(0, t) = I(0, t)\,R_L. \tag{7.105}$$

It follows, from the three previous equations, that

$$I_r = \left(\frac{Z - R_L}{Z + R_L}\right) I_i. \tag{7.106}$$

Hence, the coefficient of reflection, which is the ratio of the power reflected by the load to the power sent down the line, is

$$R = \left(\frac{I_r}{I_i}\right)^2 = \left(\frac{Z - R_L}{Z + R_L}\right)^2. \tag{7.107}$$

Furthermore, the coefficient of transmission, which is the ratio of the power absorbed by the load to the power sent down the line, takes the form

$$T = 1 - R = \frac{4\,Z\,R_L}{(Z + R_L)^2}. \tag{7.108}$$

It can be seen, by comparison with Equations (7.82) and (7.83), that the load terminating the line acts just like another transmission line of impedance R_L. Moreover, it is clear that power can only be efficiently sent down a transmission line, and transferred to a terminating load, when the impedance of the line matches the effective impedance of the load (which, in this case, is the same as the resistance of the load). In other words, when $Z = R_L$ there is no reflection of the signal sent down the line (*i.e.*, $R = 0$), and all of the signal energy is therefore absorbed by the load (*i.e.*, $T = 1$). As an example, a *half-wave antenna* (*i.e.*, an antenna whose length is half the wavelength of the emitted radiation) has a characteristic impedance of 73 Ω. Hence, a transmission line used to feed energy into such an antenna should also have a characteristic impedance of $73\,\Omega$. Suppose, however, that we encounter a situation in which the impedance of a transmission line, Z_1, does not match that of its terminating load, Z_2. Can anything be done to avoid reflection of the signal sent down the line? It turns out, by analogy with the analysis presented above, that if the line is connected to the load via a short section of transmission line whose length is one quarter of the wavelength of the signal, and whose characteristic impedance is $Z_3 = \sqrt{Z_1\,Z_2}$, then there is no reflection of the signal: *i.e.*, all of the signal power is absorbed by the load. A short section of transmission line used in this manner is known as a *quarter wave transformer*.

7.7 Electromagnetic Waves

Consider a *plane electromagnetic wave* propagating through a vacuum in the z-direction. Electromagnetic waves are, incidentally, the only commonly occurring waves which do not require a medium through which to propagate. Suppose that the wave is *polarized in the* x*-direction*: *i.e.*, its electric component oscillates in the x-direction. It follows that the magnetic component of the wave oscillates in the y-direction. According to standard electromagnetic theory, the wave is described by the following pair of coupled partial differential equations:

$$\frac{\partial E_x}{\partial t} = -\frac{1}{\epsilon_0}\,\frac{\partial H_y}{\partial z}, \tag{7.109}$$

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0}\,\frac{\partial E_x}{\partial z}, \tag{7.110}$$

where $E_x(z,t)$ is the *electric field-strength*, and $H_y(z,t)$ is the *magnetic intensity* (*i.e.*, the magnetic field-strength divided by μ_0). Observe that Equations (7.109) and (7.110), which govern the propagation of electromagnetic waves through a vacuum, are analogous to Equations (7.53) and (7.54), which govern the propagation of electromagnetic signals down a transmission line. In particular, E_x has units of voltage over length, H_y has units of current over length, ϵ_0 has units of capacitance per unit length, and μ_0 has units of inductance per unit length.

Equations (7.109) and (7.110) can be combined to give

$$\frac{\partial^2 E_x}{\partial t^2} = \frac{1}{\epsilon_0\,\mu_0}\,\frac{\partial^2 E_x}{\partial z^2}, \tag{7.111}$$

$$\frac{\partial^2 H_y}{\partial t^2} = \frac{1}{\epsilon_0\,\mu_0}\,\frac{\partial^2 H_y}{\partial z^2}. \tag{7.112}$$

It follows that the electric and the magnetic components of an electromagnetic wave propagating through a vacuum both separately satisfy a wave equation of the form (7.1). Furthermore, the phase velocity of the wave is clearly

$$c = \frac{1}{\sqrt{\epsilon_0\,\mu_0}} = 2.998\times 10^8\,\mathrm{m\,s^{-1}}. \tag{7.113}$$

Let us search for a traveling wave solution of (7.109) and (7.110), propagating in the positive z-direction, whose electric component has the form

$$E_x(z,t) = E_0\,\cos(k\,z - \omega\,t - \phi). \tag{7.114}$$

As is easily demonstrated, this is a valid solution provided that $\omega = k\,c$. According to (7.109), the magnetic component of the wave is written

$$H_y(z,t) = Z^{-1}\,E_0\,\cos(k\,z - \omega\,t - \phi), \tag{7.115}$$

where

$$Z = Z_0 \equiv \sqrt{\frac{\mu_0}{\epsilon_0}}, \tag{7.116}$$

and Z_0 is the *impedance of free space* [see Equation (7.71)]. Thus, the electric and magnetic components of an electromagnetic wave propagating through a vacuum are *mutually perpendicular*, and also *perpendicular to the direction of propagation*. Moreover, the two components oscillate *in phase* (*i.e*, they have simultaneous maxima and zeros), and the amplitude of the magnetic component is that of the electric component divided by the impedance of free space.

Multiplying (7.109) by $\epsilon_0\,E_x$, (7.110) by $\mu_0\,H_y$, and adding the two resulting expressions, we obtain the energy conservation equation

$$\frac{\partial\mathcal{E}}{\partial t} + \frac{\partial\mathcal{I}}{\partial z} = 0, \tag{7.117}$$

where

$$\mathcal{E} = \frac{1}{2}\left(\epsilon_0\,E_x^{\,2} + \mu_0\,H_y^{\,2}\right) \tag{7.118}$$

is the *electromagnetic energy per unit volume* of the wave, whereas

$$\mathcal{I} = E_x\,H_y \tag{7.119}$$

is the wave *electromagnetic energy flux* (*i.e.*, power per unit area) in the positive z-direction. The mean energy flux associated with the z-directed electromagnetic wave specified in Equations (7.114) and (7.115) is thus

$$\langle\mathcal{I}\rangle = \frac{1}{2}\,\frac{E_0^{\,2}}{Z}. \tag{7.120}$$

For a similar wave propagating in the negative z-direction, it is easily demonstrated that

$$E_x(z,t) = E_0 \cos(k\,z + \omega\,t - \phi), \tag{7.121}$$

$$H_y(z,t) = -Z^{-1}\,E_0 \cos(k\,z + \omega\,t - \phi), \tag{7.122}$$

and

$$\langle \mathcal{I} \rangle = -\frac{1}{2}\,\frac{E_0^2}{Z}. \tag{7.123}$$

Consider a plane electromagnetic wave, polarized in the x-direction, which propagates in the z-direction through a *transparent dielectric medium*, such as glass or water. As is well known, the electric component of the wave causes the neutral molecules making up the medium to *polarize*: *i.e.*, it causes a small separation to develop between the mean positions of the positively and negatively charged constituents of the molecules (*i.e.*, the atomic nuclii and the electrons). (Incidentally, it is easily shown that the magnetic component of the wave has a negligible influence on the molecules, provided that the wave amplitude is sufficiently small that the wave electric field does not cause the electrons and nuclii to move with relativistic velocities.) Now, if the mean position of the positively charged constituents of a given molecule, of net charge $+q$, develops a vector displacement $\mathbf{d}$ with respect to the mean position of the negatively charged constituents, of net charge $-q$, in response to a wave electric field $\mathbf{E}$ then the associated *electric dipole moment* is $\mathbf{p} = q\,\mathbf{d}$, where $\mathbf{d}$ is generally parallel to $\mathbf{E}$. Furthermore, if there are N such molecules per unit volume then the *dipole moment per unit volume* is written $\mathbf{P} = N\,q\,\mathbf{d}$. Now, in a conventional dielectric medium,

$$\mathbf{P} = \epsilon_0\,(\epsilon - 1)\,\mathbf{E}, \tag{7.124}$$

where $\epsilon > 1$ is a dimensionless quantity, known as the *relative dielectric constant*, which is a property of the medium in question. In the presence of a dielectric medium, Equations (7.109) and (7.110) generalize to give

$$\frac{\partial E_x}{\partial t} = -\frac{1}{\epsilon_0}\left(\frac{\partial P_x}{\partial t} + \frac{\partial H_y}{\partial z}\right), \tag{7.125}$$

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0}\,\frac{\partial E_x}{\partial z}. \tag{7.126}$$

When combined with Equation (7.124), these expressions yield

$$\frac{\partial E_x}{\partial t} = -\frac{1}{\epsilon\,\epsilon_0}\,\frac{\partial H_y}{\partial z}, \tag{7.127}$$

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0}\,\frac{\partial E_x}{\partial z}. \tag{7.128}$$

It can be seen that the above equations are just like the corresponding vacuum equations, (7.109) and (7.110), except that ϵ_0 has been replaced by $\epsilon\,\epsilon_0$. It immediately follows that the *phase velocity* of an electromagnetic wave propagating through a dielectric medium is

$$v = \frac{1}{\sqrt{\epsilon\,\epsilon_0\,\mu_0}} = \frac{c}{n}, \tag{7.129}$$

where $c = 1/\sqrt{\epsilon_0\,\mu_0}$ is the velocity of light in vacuum, and the quantity

$$n = \sqrt{\epsilon} \tag{7.130}$$

is known as the *refractive index* of the medium. Thus, an electromagnetic wave propagating through a transparent dielectric medium does so at a phase velocity which is *less* than the velocity of light in vacuum by a factor n (where $n > 1$). Furthermore, the *impedance* of a transparent dielectric medium becomes

$$Z = \sqrt{\frac{\mu_0}{\epsilon\,\epsilon_0}} = \frac{Z_0}{n}, \tag{7.131}$$

where Z_0 is the impedance of free space.

Suppose that the plane $z = 0$ forms the boundary between two transparent dielectric media of refractive indices n_1 and n_2. Let the first medium occupy the region $z < 0$, and the second the region $z > 0$. Suppose that a plane electromagnetic wave, polarized in the x-direction, and propagating in the positive z-direction, is launched toward the boundary from a wave source of angular frequency ω situated at $z = -\infty$. Of course, we expect the wave incident on the boundary to be partly reflected, and partly transmitted. The wave electric and magnetic fields in the region $z < 0$ are written

$$E_x(z,t) = E_i\,\cos(k_1\,z - \omega\,t) + E_r\,\cos(k_1\,z + \omega\,t), \tag{7.132}$$

$$H_y(z,t) = Z_1^{-1}\,E_i\,\cos(k_1\,z - \omega\,t) - Z_1^{-1}\,E_r\,\cos(k_1\,z + \omega\,t), \tag{7.133}$$

where E_i is the amplitude of (the electric component of) the incident wave, E_r the amplitude of the reflected wave, $k_1 = n_1\,\omega/c$, and $Z_1 = Z_0/n_1$. The wave electric and magnetic fields in the region $z > 0$ take the form

$$E_x(z,t) = E_t\,\cos(k_2\,z - \omega\,t), \tag{7.134}$$

$$H_y(z,t) = Z_2^{-1}\,E_t\,\cos(k_2\,z - \omega\,t), \tag{7.135}$$

where E_t is the amplitude of the transmitted wave, $k_2 = n_2\,\omega/c$, and $Z_2 = Z_0/n_2$. According to standard electromagnetic theory, the appropriate matching conditions at the boundary ($z = 0$) are simply that E_x and H_y both be *continuous*. Thus, continuity of E_x yields

$$E_i + E_r = E_t, \tag{7.136}$$

whereas continuity of H_y gives

$$n_1\,(E_i - E_r) = n_2\,E_t, \tag{7.137}$$

since $Z^{-1} \propto n$. It follows that

$$E_r = \left(\frac{n_1 - n_2}{n_1 + n_2}\right) E_i, \tag{7.138}$$

$$E_t = \left(\frac{2\,n_1}{n_1 + n_2}\right) E_i. \tag{7.139}$$

The coefficient of reflection, R, is defined as the ratio of the reflected to the incident energy flux, so that

$$R = \left(\frac{E_r}{E_i}\right)^2 = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2. \tag{7.140}$$

Likewise, the coefficient of transmission, T, is the ratio of the transmitted to the incident energy flux, so that

$$T = \frac{Z_2^{-1}}{Z_1^{-1}}\left(\frac{E_t}{E_i}\right)^2 = \frac{n_2}{n_1}\left(\frac{E_t}{E_i}\right)^2 = \frac{4\,n_1\,n_2}{(n_1 + n_2)^2} = 1 - R. \tag{7.141}$$

It can be seen, first of all, that if $n_1 = n_2$ then $E_r = 0$ and $E_t = E_i$. In other words, if the two media have the same indices of refraction then there is no reflection at the boundary between them, and the transmitted wave is consequently equal in amplitude to the incident wave. On the other hand, if $n_1 \neq n_2$ then there is always some reflection at the boundary. Indeed, the amplitude of the reflected wave is roughly proportional to the difference between n_1 and n_2. This has important practical consequences. We can only see a clean pane of glass in a window because some of the light incident on an air/glass boundary is reflected, due to the different refractive indicies of air and glass. As is well known, it is a lot more difficult to see glass when it is submerged in water. This is because the refractive indices of glass and water are quite similar, and so there is very little reflection of light incident on a water/glass boundary.

According to Equation (7.138), $E_r/E_i < 0$ when $n_2 > n_1$. The negative sign indicates a π radian phase shift of the (electric component of the) reflected wave, with respect to the incident wave. We conclude that there is a

π radian phase shift of the reflected wave, relative to the incident wave, on reflection from a boundary with a medium of *greater* refractive index. Conversely, there is no phase shift on reflection from a boundary with a medium of *lesser* refractive index.

Note that Equations (7.138)–(7.141) are analogous to Equations (7.80)–(7.83), with refractive index playing the role of impedance. This suggests, by analogy with earlier analysis, that we can prevent reflection of an electromagnetic wave normally incident at a boundary between two transparent dielectric media of different refractive indices by separating the media by a thin transparent layer whose thickness is one quarter of a wavelength, and whose refractive index is the geometric mean of the refractive indices of the two media. This is the physical principle behind the *non-reflective lens coatings* used in high-quality optical instruments.

7.8 Exercises

1. Write the traveling wave $\psi(x,t) = A\,\cos(k\,x - \omega\,t)$ as a superposition of two standing waves. Write the standing wave $\psi(x,t) = A\,\cos(k\,x)\,\cos(\omega\,t)$ as a superposition of two traveling waves propagating in opposite directions. Show that the following superposition of traveling waves,

$$\psi(x,t) = A\,\cos(k\,x - \omega\,t) + A\,R\,\cos(k\,x + \omega\,t),$$

can be written as the following superposition of standing waves,

$$\psi(x,t) = A\,(1+R)\,\cos(k\,x)\,\cos(\omega\,t) + A\,(1-R)\,\sin(k\,x)\,\sin(\omega\,t).$$

2. Demonstrate that for a transverse traveling wave propagating on a stretched string

$$\langle\mathcal{I}\rangle = v\,\langle\mathcal{E}\rangle,$$

where $\langle\mathcal{I}\rangle$ is the mean energy flux along the string due to the wave, $\langle\mathcal{E}\rangle$ is the mean wave energy per unit length, and v is the phase velocity of the wave. Show that the same relation holds for a longitudinal traveling wave in an elastic solid.

3. A transmission line of characteristic impedance Z occupies the region $x < 0$, and is terminated at $x = 0$. Suppose that the current carried by the line takes the form

$$I(x,t) = I_i\,\cos(k\,x - \omega\,t) + I_r\,\cos(k\,x + \omega\,t)$$

for $x \leq 0$, where I_i is the amplitude of the incident signal, and I_r the amplitude of the signal reflected at the end of the line. Let the end of the line be *open circuited*, such that the line is effectively terminated by an infinite resistance. Find the relationship between I_r and I_i. Show that the current and voltage oscillate $\pi/2$ radians out of phase everywhere along the line. Demonstrate that there is zero net flux of electromagnetic energy along the line.

4. Suppose that the transmission line in the previous exercise is *short circuited*, such that the line is effectively terminated by a negligible resistance. Find the relationship between I_r and I_i. Show that the current and voltage oscillate $\pi/2$ radians out of phase everywhere along the line. Demonstrate that there is zero net flux of electromagnetic energy along the line.

5. Suppose that the transmission line of Exercise 3 is terminated by an inductor of inductance L, such that

$$V(0,t) = L\,\frac{\partial I(0,t)}{\partial t}.$$

   Find the relationship between I_r and I_i. Obtain expressions for the current, $I(x,t)$, and the voltage, $V(x,t)$, along the line (which only involve I_i). Demonstrate that the incident and the reflected wave both have zero net associated energy flux.

6. Suppose that the transmission line of Exercise 3 is terminated by a capacitor of capacitance C. Find the relationship between I_r and I_i. Obtain expressions for the current, $I(x,t)$, and the voltage, $V(x,t)$, along the line (which only involve I_i). Demonstrate that the incident and the reflected wave both have zero net associated energy flux.

7. A lossy transmission line has a resistance per unit length $\mathcal{R}$, in addition to an inductance per unit length $\mathcal{L}$, and a capacitance per unit length $\mathcal{C}$. The resistance can be considered to be in series with the inductance. Demonstrate that the Telegrapher's equations generalize to

$$\begin{aligned}\frac{\partial V}{\partial t} &= -\frac{1}{\mathcal{C}}\frac{\partial I}{\partial x},\\ \frac{\partial I}{\partial t} &= -\frac{\mathcal{R}}{\mathcal{L}}\,I - \frac{1}{\mathcal{L}}\frac{\partial V}{\partial x},\end{aligned}$$

   where $I(x,t)$ and $V(x,t)$ are the voltage and current along the line. Derive an energy conservation equation of the form

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial \mathcal{I}}{\partial x} = -\mathcal{R}\,I^2,$$

   where $\mathcal{E}$ is the energy per unit length along the line, and $\mathcal{I}$ the energy flux. Give expressions for $\mathcal{E}$ and $\mathcal{I}$. What does the right-hand side of the above equation represent? Show that the current obeys the wave-diffusion equation

$$\frac{\partial^2 I}{\partial t^2} + \frac{\mathcal{R}}{\mathcal{L}}\frac{\partial I}{\partial t} = \frac{1}{\mathcal{L}\mathcal{C}}\frac{\partial^2 I}{\partial x^2}.$$

Consider the low resistance, high frequency limit $\omega \gg \mathcal{R}/\mathcal{L}$. Demonstrate that a signal propagating down the line varies as

$$\begin{aligned} I(x,t) &\simeq I_0 \cos[k(x - v\,t)]\,e^{-x/\delta}, \\ V(x,t) &\simeq Z\,I_0 \cos[k(x - v\,t)]\,e^{-x/\delta}, \end{aligned}$$

where $k = \omega/v$, $v = 1/\sqrt{\mathcal{L}\mathcal{C}}$, $\delta = 2\,Z/\mathcal{R}$, and $Z = \sqrt{\mathcal{L}/\mathcal{C}}$. Show that $k\,\delta \ll 1$: *i.e.*, that the decay length of the signal is much longer than its wavelength. Estimate the maximum useful length of a low resistance, high frequency, lossy transmission line.

8. Suppose that a transmission line consisting of two uniform parallel conducting strips of width w and perpendicular distance apart d, where $d \ll w$, is terminated by a strip of material of uniform resistance per square meter $\sqrt{\mu_0/\epsilon_0} = 376.73\,\Omega$. Such material is known as *spacecloth*. Demonstrate that a signal sent down the line is completely absorbed, with no reflection, by the spacecloth. Incidentally, the resistance of a uniform strip of material is proportional to its length, and inversely proportional to its cross-sectional area.

9. At normal incidence, the mean radiant power from the Sun illuminating one square meter of the Earth's surface is $1.35\,\text{kW}$. Show that the amplitude of the electric component of solar electromagnetic radiation at the Earth's surface is $1010\,\text{V}\,\text{m}^{-1}$. Demonstrate that the corresponding amplitude of the magnetic component is $2.7\,\text{A}\,\text{m}^{-1}$.

10. According to Einstein's famous formula, $E = m\,c^2$, where E is energy, m is mass, and c is the velocity of light in vacuum. This formula implies that anything that possesses energy also has an effective mass. Use this idea to show that an electromagnetic wave of mean intensity (energy per unit time per unit area) $\langle\mathcal{I}\rangle$ has an associated mean pressure (momentum per unit time per unit area) $\langle\mathcal{P}\rangle = \langle\mathcal{E}\rangle/c$. Hence, estimate the pressure due to sunlight at the Earth's surface.

11. A glass lens is coated with a non-reflecting coating of thickness one quarter of a wavelength (in the coating) of light whose wavelength in air is λ_0. The index of refraction of the glass is n, and that of the coating is $\sqrt{n}$. The refractive index of air can be taken to be unity. Show that the coefficient of reflection for light normally incident on the lens from air is

$$R = 4\left(\frac{1-\sqrt{n}}{1+\sqrt{n}}\right)^2 \sin^2\left(\frac{\pi}{2}\left[\frac{\lambda_0}{\lambda} - 1\right]\right),$$

where λ is the wavelength of the incident light in air. Assume that $n = 1.5$, and that this value remains approximately constant for light whose wavelengths lie in the visible band. Suppose that $\lambda_0 = 550\,\text{nm}$, which corresponds to green light. It follows that $R = 0$ for green light. What is R for blue light of wavelength $\lambda = 450\,\text{nm}$, and for red light of wavelength $650\,\text{nm}$? Comment on how effective the coating is at suppressing unwanted reflection of visible light incident on the lens.

12. A glass lens is coated with a non-reflective coating whose thickness is one quarter of a wavelength (in the coating) of light whose frequency is f_0. Demonstrate that the coating also suppresses reflection from light whose frequency is $3\,f_0$, $5\,f_0$, *etc.*, assuming that the refractive index of the coating and the glass is frequency independent.

13. An plane electromagnetic wave, polarized in the x-direction, and propagating in the z-direction though a conducting medium of conductivity σ is governed by

$$\begin{aligned}\frac{\partial E_x}{\partial t} &= -\frac{\sigma}{\epsilon_0}\,E_x - \frac{1}{\epsilon_0}\,\frac{\partial H_y}{\partial z},\\ \frac{\partial H_y}{\partial t} &= -\frac{1}{\mu_0}\,\frac{\partial E_x}{\partial z},\end{aligned}$$

where $E_x(z,t)$ and $H_y(z,t)$ are the electric and magnetic components of the wave. Derive an energy conservation equation of the form

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\partial \mathcal{I}}{\partial x} = -\sigma\,E_x^{\,2},$$

where $\mathcal{E}$ is the electromagnetic energy per unit volume, and $\mathcal{I}$ the electromagnetic energy flux. Give expressions for $\mathcal{E}$ and $\mathcal{I}$. What does the right-hand side of the above equation represent? Demonstrate that E_x obeys the wave-diffusion equation

$$\frac{\partial^2 E_x}{\partial t^2} + \frac{\sigma}{\epsilon_0}\,\frac{\partial E_x}{\partial t} = c^2\,\frac{\partial^2 E_x}{\partial z^2},$$

where $c = 1/\sqrt{\epsilon_0\,\mu_0}$. In the high frequency, low conductivity limit $\omega \gg \sigma/\epsilon_0$, show that the above equation has the approximate solution

$$E_x(z,t) \simeq E_0\,\cos[k\,(z - c\,t)]\,e^{-z/\delta},$$

where $k = \omega/c$, $\delta = 2/(Z_0\,\sigma)$, and $Z_0 = \sqrt{\mu_0/\epsilon_0}$. What is the corresponding solution for $H_y(z,t)$? Demonstrate that $k\,\delta \ll 1$: *i.e.*, that the wave penetrates many wavelengths into the medium. Estimate how far a high frequency electromagnetic wave penetrates into a low conductivity conducting medium.

8 Wave Pulses

8.1 Fourier Transforms

Consider a function $F(x)$ which is *periodic* in x with period L. In other words,

$$F(x+L) = F(x) \tag{8.1}$$

for all x. Recall, from Section 6.4, that we can represent such a function as a *Fourier series*: *i.e.*,

$$F(x) = \sum_{n=1,\infty} \left[C_n \cos(n\,\delta k\,x) + S_n \sin(n\,\delta k\,x)\right], \tag{8.2}$$

where

$$\delta k = \frac{2\pi}{L}. \tag{8.3}$$

[Note that we have neglected the $n = 0$ term in (8.2), for the sake of convenience.] The expression (8.2) *automatically* satisfies the periodicity constraint (8.1), since $\cos(\theta + n\,2\pi) = \cos\theta$ and $\sin(\theta + n\,2\pi) = \sin\theta$ for all θ and n (with the proviso that n is integer). The so-called *Fourier coefficients*, C_n and S_n, appearing in (8.2), can be determined from the function $F(x)$ by means of the following easily demonstrated results (see Exercise 1):

$$\frac{2}{L}\int_{-L/2}^{L/2} \cos(n\,\delta k\,x)\,\cos(n'\,\delta k\,x)\,dx = \delta_{n,n'}, \tag{8.4}$$

$$\frac{2}{L}\int_{-L/2}^{L/2} \sin(n\,\delta k\,x)\,\sin(n'\,\delta k\,x)\,dx = \delta_{n,n'}, \tag{8.5}$$

$$\frac{2}{L}\int_{-L/2}^{L/2} \cos(n\,\delta k\,x)\,\sin(n'\,\delta k\,x)\,dx = 0, \tag{8.6}$$

where n, n' are positive integers. Here, $\delta_{n,n'} = 1$ if $n = n'$, and $\delta_{n,n'} = 0$ otherwise. In fact (see Exercise 1),

$$C_n = \frac{2}{L}\int_{-L/2}^{L/2} F(x)\,\cos(n\,\delta k\,x)\,dx, \tag{8.7}$$

$$S_n = \frac{2}{L}\int_{-L/2}^{L/2} F(x)\,\sin(n\,\delta k\,x)\,dx. \tag{8.8}$$

Note, finally, that *any* periodic function of x can be represented as a Fourier series.

Suppose, however, that we are dealing with a function $F(x)$ which is *not* periodic in x. Actually, we can think of such a function as one which is periodic in x with a period L that tends to infinity. Does this mean that we can still represent $F(x)$ as a Fourier series? Consider what happens to the series (8.2) in the limit that $L \to \infty$, or, equivalently, $\delta k \to 0$. Now, the series is basically a weighted sum of sinusoidal functions whose wavenumbers take the *quantized* values $k_n = n\,\delta k$. Moreover, as $\delta k \to 0$, these values become more and more closely spaced. In fact, we can write

$$F(x) = \sum_{n=1,\infty} \frac{C_n}{\delta k} \cos(n\,\delta k\,x)\,\delta k + \sum_{n=1,\infty} \frac{S_n}{\delta k} \sin(n\,\delta k\,x)\,\delta k. \tag{8.9}$$

In the continuum limit, $\delta k \to 0$, the *summations* in the above expression become *integrals*, and we obtain

$$F(x) = \int_{-\infty}^{\infty} C(k)\,\cos(k\,x)\,dk + \int_{-\infty}^{\infty} S(k)\,\sin(k\,x)\,dk, \tag{8.10}$$

where $k = n\,\delta k$, $C(k) = C(-k) = C_n/(2\,\delta k)$, and $S(k) = -S(-k) = S_n/(2\,\delta k)$. Thus, for the case of an aperiodic function, the *Fourier series* (8.2) morphs into the so-called *Fourier transform* (8.10). This transform can be inverted using the continuum limit (*i.e.*, the limit $\delta k \to 0$) of Equations (8.7) and (8.8), which are easily be shown to be

$$C(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(x)\,\cos(k\,x)\,dx, \tag{8.11}$$

$$S(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(x)\,\sin(k\,x)\,dx. \tag{8.12}$$

Incidentally, it is clear, from the above equations, that $C(-k) = C(k)$ and $S(-k) = -S(k)$. The *Fourier space* (*i.e.*, k-space) functions $C(k)$ and $S(k)$ are known as the *cosine Fourier transform* and the *sine Fourier transform* of the *real space* (*i.e.*, x-space) function $F(x)$, respectively. Furthermore, since we already know that any periodic function can be represented as a Fourier series, it seems plausible that any aperiodic function can be represented as a Fourier transform. This is indeed the case. Note that Equations (8.10)–(8.12) effectively enable us to represent a general function as a linear superposition of sine and cosine functions. Let us now consider some examples.

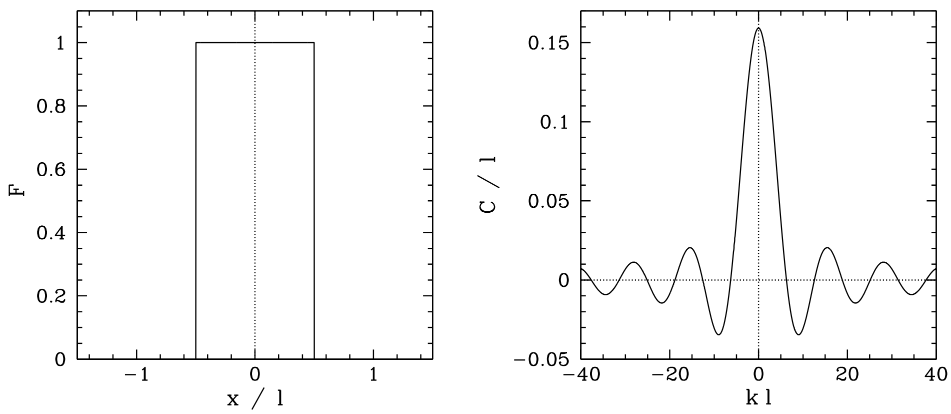


Figure 8.1: *Fourier transform of a square-wave function.*

Consider the *square-wave* function (see Figure 8.1)

$$F(x) = \begin{cases} 1 & |x| \leq l/2 \\ 0 & |x| > l/2 \end{cases}. \tag{8.13}$$

Now, given that $\cos(-k\,x) = \cos(k\,x)$ and $\sin(-k\,x) = -\sin(k\,x)$, it is apparent, from (8.11) and (8.12), that if $F(x)$ is *even* in x, so that $F(-x) = F(x)$, then $S(k) = 0$, and if $F(x)$ is *odd* in x, so that $F(-x) = -F(x)$, then $C(k) = 0$. Hence, since the square-wave function (8.13) is clearly even in x, its sine Fourier transform is automatically zero. On the other hand, its cosine Fourier transform takes the form

$$C(k) = \frac{1}{2\pi}\int_{-l/2}^{l/2} \cos(k\,x)\,dx = \frac{l}{2\pi}\,\frac{\sin(k\,l/2)}{k\,l/2}. \tag{8.14}$$

Figure 8.1 shows the function $F(x)$, together with its associated cosine transform, $C(k)$.

As a second example, consider the *Gaussian* function

$$F(x) = \exp\left(-\frac{x^2}{2\,\sigma_x^2}\right). \tag{8.15}$$

As illustrated in Figure 8.2, this is a smoothly varying even function of x which attains its peak value 1 at $x = 0$, and becomes completely negligible when $|x| \gtrsim 3\,\sigma_x$. Thus, σ_x is a measure of the "width" of the function in real space. By symmetry, the sine Fourier transform of the above function is

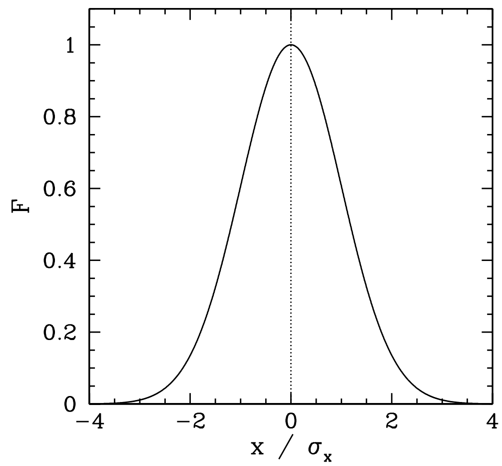


Figure 8.2: *A Gaussian function.*

zero. On the other hand, the cosine Fourier transform is easily shown to be (see Exercise 2)

$$C(k) = \frac{1}{(2\pi\,\sigma_k^2)^{1/2}}\,\exp\left(-\frac{k^2}{2\,\sigma_k^2}\right), \tag{8.16}$$

where

$$\sigma_k = \frac{1}{\sigma_x}. \tag{8.17}$$

Note that this function is a *Gaussian* in Fourier space of characteristic width $\sigma_k = 1/\sigma_x$. In fact, the Gaussian is the only mathematical function which is its own Fourier transform. Now, the original function $F(x)$ can be reconstructed from its Fourier transform using

$$F(x) = \int_{-\infty}^{\infty} C(k)\,\cos(k\,x)\,dk. \tag{8.18}$$

This reconstruction is simply a linear superposition of cosine waves of differing wavenumbers. Moreover, $C(k)$ can be interpreted as the amplitude of waves of wavenumber k within this superposition. The fact that $C(k)$ is a Gaussian of characteristic width $\sigma_k = 1/\sigma_x$ [which means that $C(k)$ is negligible for $|k| \gtrsim 3\,\sigma_k$] implies that in order to reconstruct a real space function whose width in real space is approximately σ_x it is necessary to combine sinusoidal functions with a range of different wavenumbers which is approximately $\sigma_k = 1/\sigma_x$ in extent. To be slightly more exact, the real-space Gaussian function $F(x)$ falls to *half* of its peak value when $|x| \simeq \sqrt{\pi/2}\,\sigma_x$. Hence, the *full width at half maximum* of the function is $\Delta x = 2\sqrt{\pi/2}\,\sigma_x = \sqrt{2\pi}\,\sigma_x$. Likewise, the full width at half maximum of the Fourier space Gaussian function $C(k)$ is $\Delta k = \sqrt{2\,\pi}\,\sigma_k$. Thus,

$$\Delta x\,\Delta k = 2\pi, \tag{8.19}$$

since $\sigma_k\,\sigma_x = 1$. Thus, a function which is highly localized in real space has a transform which is highly delocalized in Fourier space, and *vice versa*. Note, finally, that (see Exercise 3)

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi\,\sigma_k^2)^{1/2}}\,\exp\left(-\frac{k^2}{2\,\sigma_k^2}\right) dk = 1. \tag{8.20}$$

In other words, a Gaussian function in real space, of *unit* height and characteristic width σ_x, has a cosine Fourier transform which is a Gaussian in Fourier space, of characteristic width $\sigma_k = 1/\sigma_x$, whose integral over all k-space is *unity*.

Consider what happens to the above mentioned real space Gaussian, and its Fourier transform, in the limit $\sigma_x \to \infty$, or, equivalently, $\sigma_k \to 0$. There is no difficulty in seeing, from Equation (8.15), that

$$F(x) \to 1. \tag{8.21}$$

In other words, the real space Gaussian morphs into a function which takes the constant value unity everywhere. The Fourier transform is more problematic. In the limit $\sigma_k \to 0$, Equation (8.16) yields a k-space function which is zero everywhere apart from $k = 0$ (since the function is negligible for $|k| \gtrsim \sigma_k$), where it is infinite [since the function takes the value $(2\pi\,\sigma_k)^{-1/2}$ at $k = 0$]. Moreover, according to Equation (8.20), the integral of the function over all k remains unity. Thus, the Fourier transform of the uniform function $F(x) = 1$ is a sort of integrable "spike" located at $k = 0$. This strange function is known as the *Dirac delta function*, and is denoted $\delta(k)$. Thus, one definition of a delta function is

$$\delta(k) \equiv \lim_{\sigma_k \to 0} \frac{1}{(2\pi\,\sigma_k^2)^{1/2}}\,\exp\left(-\frac{k^2}{2\,\sigma_k^2}\right). \tag{8.22}$$

As has already been mentioned, $\delta(k) = 0$ for $k \neq 0$, and $\delta(0) = \infty$. Moreover,

$$\int_{-\infty}^{\infty} \delta(k)\,dk = 1. \tag{8.23}$$

Consider the integral

$$\int_{-\infty}^{\infty} F(k)\,\delta(k)\,dk, \tag{8.24}$$

where $F(k)$ is an arbitrary function. Because of the peculiar properties of the delta function, the only contribution to the above integral comes from the region in k-space in the immediate vicinity of $k = 0$. Furthermore, provided $F(k)$ is well-behaved in this region, we can write

$$\int_{-\infty}^{\infty} F(k)\,\delta(k)\,dk = \int_{-\infty}^{\infty} F(0)\,\delta(k)\,dk = F(0)\int_{-\infty}^{\infty} \delta(k)\,dk = F(0), \tag{8.25}$$

where use has been made of Equation (8.23).

A simple change of variables allows us to define $\delta(k-k')$, which is a "spike" function centered on $k = k'$. The above result can easily be generalized to give

$$\int_{-\infty}^{\infty} F(k)\,\delta(k-k')\,dk = F(k'), \tag{8.26}$$

for all $F(k)$. Indeed, this expression can be thought of as an alternative definition of a delta function.

Now, we have seen that the delta function $\delta(k)$ is the cosine Fourier transform of the uniform function $F(x) = 1$. It, thus, follows from (8.11) that

$$\delta(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos(k\,x)\,dx. \tag{8.27}$$

This result represents yet another definition of the delta function. By symmetry, we also have

$$0 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \sin(k\,x)\,dx. \tag{8.28}$$

It follows that

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} \cos(k\,x)\,\cos(k'\,x)\,dx &= \frac{1}{4\pi}\int_{-\infty}^{\infty} \left\{\cos\left[(k-k')\,x\right] + \right.\\ &\qquad \left.\cos\left[(k+k')\,x\right]\right\}dx\\ &= \frac{1}{2}\left[\delta(k-k') + \delta(k+k')\right], \end{aligned} \tag{8.29}$$

where use has been made of (8.27) and a standard trigonometric identity. Likewise (see Exercise 4),

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \sin(k\,x)\,\sin(k'\,x)\,dx = \frac{1}{2}\left[\delta(k-k') - \delta(k+k')\right], \tag{8.30}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\cos(k\,x)\,\sin(k'\,x)\,dx \;=\; 0. \tag{8.31}$$

Incidentally, Equations (8.29)–(8.31) can be used to derive Equations (8.11)–(8.12) directly from Equation (8.10) (see Exercise 5).

8.2 General Solution of the Wave Equation

Consider the one dimensional wave equation

$$\frac{\partial^2\psi}{\partial t^2} = v^2\,\frac{\partial^2\psi}{\partial x^2}, \tag{8.32}$$

where $\psi(x,t)$ is the wave amplitude, and v the characteristic phase velocity. We have seen a number of particular solutions of this equation. For instance,

$$\psi(x,t) = A\,\cos(k\,x - \omega\,t - \phi) \tag{8.33}$$

represents a traveling wave of amplitude A, wavenumber k, angular frequency ω, and phase angle ϕ, which propagates in the positive x-direction. The above expression is a solution of the wave equation (8.32) provided that it satisfies the *dispersion relation*

$$\omega = k\,v: \tag{8.34}$$

i.e., provided that the wave propagates with the fixed phase velocity v. We can also write the solution (8.33) as

$$\psi(x,t) = C_+\,\cos[k\,(x - v\,t)] + S_+\,\sin[k\,(x - v\,t)], \tag{8.35}$$

where $C_+ = A\,\cos\phi$, $S_+ = A\,\sin\phi$, and we have explicitly incorporated the dispersion relation $\omega = k\,v$ into the solution. The above expression can be regarded as the most general form for a traveling wave of wavenumber k propagating in the positive x-direction. Likewise, the most general form for a traveling wave of wavenumber k propagating in the negative x-direction is

$$\psi(x,t) = C_-\,\cos[k\,(x + v\,t)] + S_-\,\sin[k\,(x + v\,t)]. \tag{8.36}$$

Of course, we have also encountered standing wave solutions of (8.32). However, as we have seen, these can be regarded as linear superpositions of traveling waves, of equal amplitude and wavenumber, propagating in opposite directions. In other words, standing waves are not fundamentally different to traveling waves.

Now, the wave equation (8.32) is *linear*. This suggests that its most general solution can be written as a *linear superposition* of all of its valid wavelike solutions. In the absence of specific boundary conditions, there is no restriction on the possible wavenumbers of such solutions. Thus, it is plausible that the most general solution of (8.32) can be written

$$\begin{aligned}\psi(x,t) &= \int_{-\infty}^{\infty} C_+(k)\,\cos[k\,(x-v\,t)]\,dk \\ &\quad+\int_{-\infty}^{\infty} S_+(k)\,\sin[k\,(x-v\,t)]\,dk \\ &\quad+\int_{-\infty}^{\infty} C_-(k)\,\cos[k\,(x+v\,t)]\,dk \\ &\quad+\int_{-\infty}^{\infty} S_-(k)\,\sin[k\,(x+v\,t)]\,dk: \qquad (8.37)\end{aligned}$$

i.e., as a linear superposition of traveling waves propagating to the right (*i.e.*, in the positive x-direction) and to the left. Here, $C_+(k)$ represents the amplitude of right-propagating cosine waves of wavenumber k in this superposition. Moreover, $S_+(k)$ represents the amplitude of right-propagating sine waves of wavenumber k, $C_-(k)$ the amplitude of left-propagating cosine waves, and $S_-(k)$ the amplitude of left-propagating sine waves. Since each of these waves is individually a solution of (8.32), we are guaranteed, from the linear nature of this equation, that the above superposition is also a solution.

But, how can we prove that (8.37) is the *most general* solution of the wave equation (8.32)? Well, our understanding of Newtonian dynamics tells us that if we know the initial wave amplitude $\psi(x,0)$, and its time derivative $\dot{\psi}(x,0)$, then this should constitute sufficient information to uniquely specify the solution of (8.32) at all subsequent times. Hence, if (8.37) is the most general solution of (8.32) then it must be consistent with *any* initial wave amplitude, and *any* initial wave velocity. In other words, given any $\psi(x,0)$ and $\dot{\psi}(x,0)$, we should be able to uniquely determine the functions $C_+(k)$, $S_+(k)$, $C_-(k)$, and $S_-(k)$ appearing in (8.37). Let us see if this is the case.

Now, from (8.37),

$$\begin{aligned}\psi(x,0) &= \int_{-\infty}^{\infty} [C_+(k)+C_-(k)]\,\cos(k\,x)\,dk \\ &\quad+\int_{-\infty}^{\infty} [S_+(k)+S_-(k)]\,\sin(k\,x)\,dk. \qquad (8.38)\end{aligned}$$

However, this is just a Fourier transform of the form (8.10). Moreover, Equations (8.11) and (8.12) allow us to uniquely invert this transform. In fact,

$$C_+(k) + C_-(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi(x,0)\,\cos(k\,x)\,dx, \tag{8.39}$$

$$S_+(k) + S_-(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi(x,0)\,\sin(k\,x)\,dx. \tag{8.40}$$

Equation (8.37) also yields

$$\begin{aligned}\dot{\psi}(x,0) &= \int_{-\infty}^{\infty} k\,v\,[C_+(k) - C_-(k)]\,\sin(k\,x)\,dk \\ &\quad - \int_{-\infty}^{\infty} k\,v\,[S_+(k) - S_-(k)]\,\cos(k\,x)\,dk.\end{aligned} \tag{8.41}$$

This is, again, a Fourier transform which can be inverted to give

$$k\,v\,[C_+(k) - C_-(k)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \dot{\psi}(x,0)\,\sin(k\,x)\,dx, \tag{8.42}$$

$$k\,v\,[S_-(k) - S_+(k)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \dot{\psi}(x,0)\,\cos(k\,x)\,dx. \tag{8.43}$$

Hence,

$$\begin{aligned}C_+(k) &= \frac{1}{4\pi}\left[\int_{-\infty}^{\infty} \psi(x,0)\,\cos(k\,x)\,dx + \int_{-\infty}^{\infty} \frac{\dot{\psi}(x,0)}{k\,v}\,\sin(k\,x)\,dx\right], \\ C_-(k) &= \frac{1}{4\pi}\left[\int_{-\infty}^{\infty} \psi(x,0)\,\cos(k\,x)\,dx - \int_{-\infty}^{\infty} \frac{\dot{\psi}(x,0)}{k\,v}\,\sin(k\,x)\,dx\right], \\ S_+(k) &= \frac{1}{4\pi}\left[\int_{-\infty}^{\infty} \psi(x,0)\,\sin(k\,x)\,dx - \int_{-\infty}^{\infty} \frac{\dot{\psi}(x,0)}{k\,v}\,\cos(k\,x)\,dx\right], \\ S_-(k) &= \frac{1}{4\pi}\left[\int_{-\infty}^{\infty} \psi(x,0)\,\sin(k\,x)\,dx + \int_{-\infty}^{\infty} \frac{\dot{\psi}(x,0)}{k\,v}\,\cos(k\,x)\,dx\right].\end{aligned} \tag{8.44}$$

It follows that we can indeed uniquely determine the functions $C_+(k)$, $C_-(k)$, $S_+(k)$, and $S_-(k)$, appearing in expression (8.37), for any $\psi(x,0)$ and $\dot{\psi}(x,0)$. This proves that (8.37) is the most general solution of the wave equation (8.32).

Let us examine our solution in more detail. Equation (8.37) can be written (see Exercise 6)

$$\psi(x,t) = F(x - v\,t) + G(x + v\,t), \tag{8.45}$$

where

$$F(x) = \int_{-\infty}^{\infty} [C_+(k)\,\cos(k\,x) + S_+(k)\,\sin(k\,x)]\,dk, \tag{8.46}$$

$$G(x) = \int_{-\infty}^{\infty} [C_-(k)\,\cos(k\,x) + S_-(k)\,\sin(k\,x)]\,dk. \tag{8.47}$$

What does the expression (8.45) signify? Well, $F(x - v\,t)$ represents a wave disturbance of *arbitrary shape* which propagates in the *positive* x-direction, at the fixed speed v, *without changing shape*. This should be clear, since a point with a given amplitude on the wave, $F(x - v\,t) = c$, has an equation of motion $x - v\,t = F^{-1}(c) =$ constant, and thus propagates in the positive x-direction at the velocity v. Moreover, since all points on the wave propagate in the same direction at the same velocity it follows that the wave does not change shape as it moves. Of course, $G(x + v\,t)$ represents a wave disturbance of arbitrary shape which propagates in the *negative* x-direction, at the fixed speed v, without changing shape. Thus, we conclude that the most general solution to the wave equation (8.32) is a superposition of two wave disturbances of arbitrary shapes which propagate in opposite directions, at the fixed speed v, without changing shape. Such solutions are generally termed *wave pulses*. So, what is the relationship between a general wave pulse and the sinusoidal traveling wave solutions to the wave equation that we found previously. Well, as is clear from Equations (8.46) and (8.47), a wave pulse can be thought of as a superposition of sinusoidal traveling waves propagating in the same direction as the pulse. Moreover, the amplitude of cosine waves of wavenumber k in this superposition is simply equal to the cosine Fourier transform of the pulse shape evaluated at wavenumber k. Likewise, the amplitude of sine waves of wavenumber k in the superposition is equal to the sine Fourier transform of the pulse shape evaluated at wavenumber k.

For instance, suppose that we have a triangular wave pulse of the form (see Figure 8.3)

$$F(x) = \begin{cases} 1 - 2\,|x|/l & |x| \leq l/2 \\ 0 & |x| > l/2 \end{cases}. \tag{8.48}$$

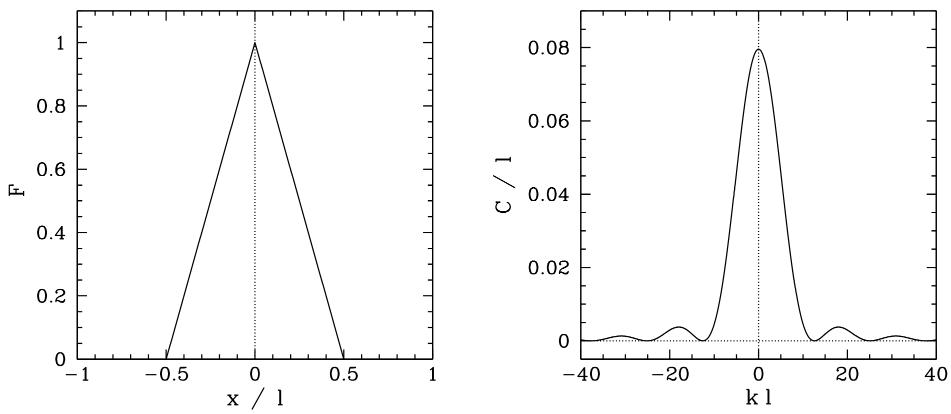


Figure 8.3: *Fourier transform of a triangular wave pulse.*

The sine Fourier transform of this pulse shape is zero by symmetry. However, the cosine Fourier transform is (see Exercise 7)

$$C(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(x)\,\cos(k\,x)\,dx = \frac{l}{4\pi}\,\frac{\sin^2(k\,l/4)}{(k\,l/4)^2}. \tag{8.49}$$

The functions $F(x)$ and $C(k)$ are shown in Figure 8.3. It follows that the right-propagating triangular wave pulse

$$\psi(x,t) = \begin{cases} 1 - 2\,|x - v\,t|/l & |x - v\,t| \leq l/2 \\ 0 & |x - v\,t| > l/2 \end{cases} \tag{8.50}$$

can be written as the following superposition of right-propagating cosine waves:

$$\psi(x,t) = \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{\sin^2(k\,l/4)}{(k\,l/4)^2}\,\cos[k\,(x - v\,t)]\,l\,dk. \tag{8.51}$$

Likewise, the left-propagating triangular wave pulse

$$\psi(x,t) = \begin{cases} 1 - 2\,|x + v\,t|/l & |x + v\,t| \leq l/2 \\ 0 & |x + v\,t| > l/2 \end{cases} \tag{8.52}$$

becomes

$$\psi(x,t) = \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{\sin^2(k\,l/4)}{(k\,l/4)^2}\,\cos[k\,(x + v\,t)]\,l\,dk. \tag{8.53}$$

8.3 Bandwidth

It is possible to Fourier transform in time, as well as in space. Thus, a general temporal waveform $F(t)$ can be written as a superposition of sinusoidal waveforms of various angular frequencies, ω: *i.e.*,

$$F(t) = \int_{-\infty}^{\infty} C(\omega)\,\cos(\omega\,t)\,d\omega + \int_{-\infty}^{\infty} S(\omega)\,\sin(\omega\,t)\,d\omega, \tag{8.54}$$

where $C(\omega)$ and $S(\omega)$ are the temporal cosine and sine Fourier transforms of the waveform, respectively. By analogy with Equations (8.10)–(8.12), we can invert the above expression to give

$$C(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(t)\,\cos(\omega\,t)\,dt, \tag{8.55}$$

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(t)\,\sin(\omega\,t)\,dt. \tag{8.56}$$

These equations make it clear that $C(-\omega) = C(\omega)$, and $S(-\omega) = -S(\omega)$. Moreover, it is apparent that if $F(t)$ is an even function of t then $S(\omega) = 0$, but if it is an odd function then $C(\omega) = 0$.

The current in the antenna of an amplitude-modulated (AM) radio transmitter is driven by a voltage signal which oscillates sinusoidally at a frequency, ω_0, which is known as the *carrier frequency*. In fact, in commercial (medium wave) AM radio, each station is assigned a single carrier frequency which lies somewhere between about 500 kHz and 1600 kHz. However, the voltage signal fed to the antenna does not have a constant amplitude. Rather, it has a *modulated amplitude* which can be expressed, somewhat schematically, as a Fourier series:

$$A(t) = A_0 + \sum_{n>0} A_n\,\cos(\omega_n\,t - \phi_n), \tag{8.57}$$

where $A(t) - A_0$ represents the information being transmitted. Typically, this information is speech or music which is picked up by a microphone, and converted into an electrical signal. Note that the constant amplitude A_0 is present even when the transmitter is transmitting no information. The remaining terms in the above expression are due to the signal picked up by the microphone. The modulation frequencies, ω_n, are thus the frequencies of audible sound waves: *i.e.*, they are so-called *audio frequencies* lying between about 20 Hz and 20 kHz. Of course, this implies that the modulation frequencies are much smaller than the carrier frequency: *i.e.*, $\omega_n \ll \omega_0$ for all $n > 0$. Furthermore, the modulation amplitudes A_n are all generally smaller than the carrier amplitude A_0.

The signal transmitted by an AM station, and received by an AM receiver, is an amplitude modulated sinusoidal oscillation of the form

$$\begin{aligned}\psi(t) &= A(t)\,\cos(\omega_0\,t)\\ &= A_0\,\cos(\omega_0\,t)+\sum_{n>0}A_n\,\cos(\omega_n\,t-\phi_n)\,\cos(\omega_0\,t),\end{aligned}\tag{8.58}$$

which, with the help of some standard trigonometric identities, can also be written

$$\begin{aligned}\psi(t) &= A_0\,\cos(\omega_0\,t)+\frac{1}{2}\sum_{n>0}A_n\,\cos[(\omega_0+\omega_n)\,t-\phi_n)]\\ &\quad+\frac{1}{2}\sum_{n>0}A_n\,\cos[(\omega_0-\omega_n)\,t+\phi_n)]\\ &= A_0\,\cos(\omega_0\,t)+\frac{1}{2}\sum_{n>0}A_n\,\cos\phi_n\,\cos[(\omega_0+\omega_n)\,t]\\ &\quad+\frac{1}{2}\sum_{n>0}A_n\,\sin\phi_n\,\sin[(\omega_0+\omega_n)\,t]\\ &\quad+\frac{1}{2}\sum_{n>0}A_n\,\cos\phi_n\,\cos[(\omega_0-\omega_n)\,t]\\ &\quad-\frac{1}{2}\sum_{n>0}A_n\,\sin\phi_n\,\sin[(\omega_0-\omega_n)\,t].\end{aligned}\tag{8.59}$$

We can calculate the cosine and sine Fourier transforms of the signal,

$$C(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\psi(t)\,\cos(\omega\,t)\,dt,\tag{8.60}$$

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\psi(t)\,\sin(\omega\,t)\,dt,\tag{8.61}$$

by making use of the standard results [*cf.*, (8.29)–(8.31)]

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\cos(\omega\,t)\,\cos(\omega'\,t)\,dt = \frac{1}{2}\left[\delta(\omega-\omega')+\delta(\omega+\omega')\right],\tag{8.62}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\sin(\omega\,t)\,\sin(\omega'\,t)\,dt = \frac{1}{2}\left[\delta(\omega-\omega')-\delta(\omega+\omega')\right], \tag{8.63}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\cos(\omega\,t)\,\sin(\omega'\,t)\,dt = 0. \tag{8.64}$$

It thus follows that

$$C(\omega>0) = \frac{1}{2}A_0\,\delta(\omega-\omega_0) + \frac{1}{4}\sum_{n>0}A_n\,\cos\phi_n\,[\delta(\omega-\omega_0-\omega_n)+\delta(\omega-\omega_0+\omega_n)], \tag{8.65}$$

$$S(\omega>0) = \frac{1}{4}\sum_{n>0}A_n\,\sin\phi_n\,[\delta(\omega-\omega_0-\omega_n)-\delta(\omega-\omega_0+\omega_n)]. \tag{8.66}$$

Here, we have only shown the positive frequency components of $C(\omega)$ and $S(\omega)$, since we know that $C(-\omega)=C(\omega)$ and $S(-\omega)=-S(\omega)$.

The AM frequency spectrum specified in Equations (8.65) and (8.66) is shown, somewhat schematically, in Figure 8.4. The spectrum consists of a series of delta function spikes. The largest spike corresponds to the carrier frequency, ω_0. However, this spike carries *no information*. Indeed, the signal information is carried in so-called *sidebands* which are equally spaced on either side of the carrier frequency. The *upper sidebands* correspond to the frequencies $\omega_0+\omega_n$, whereas the *lower sidebands* correspond to the frequencies $\omega_0-\omega_n$. Thus, in order for an AM radio signal to carry all of the information present in audible sound, for which the appropriate modulation frequencies, ω_n, range from about 0 Hz to about 20 kHz, the signal would have to consist of a superposition of sinusoidal oscillations with frequencies which range from the carrier frequency minus 20 kHz to the carrier frequency plus 20 kHz. In other words, the signal would have to occupy a *range of frequencies* from $\omega_0-\omega_N$ to $\omega_0+\omega_N$, where ω_N is the largest modulation frequency. This is an important result. An AM radio signal which only consists of a single frequency, such as the carrier frequency, transmits no information. Only a signal which occupies a finite range of frequencies, centered on the carrier frequency, is capable of transmitting useful information. The difference between the highest and the lowest frequency components of an AM radio signal, which is twice the maximum modulation frequency, is called the *bandwidth* of the signal. Thus, to transmit all

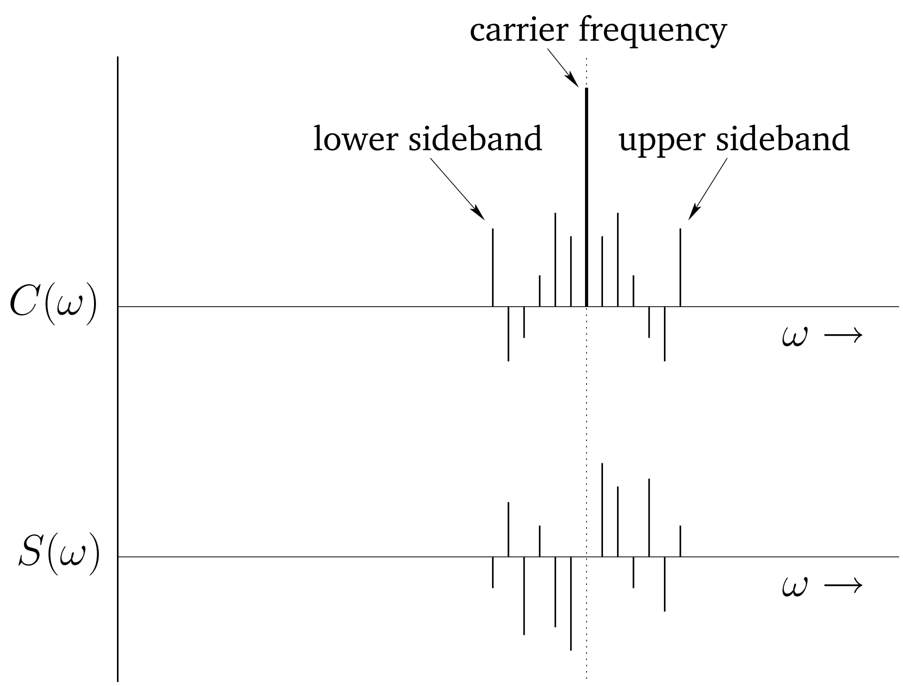


Figure 8.4: *Frequency spectrum of an AM radio signal.*

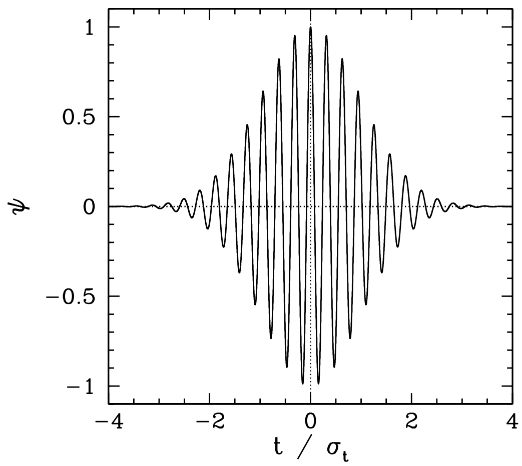


Figure 8.5: *A digital bit transmitted over AM radio.*

the information present in audible sound an AM signal would need to have a bandwidth of 40 kHz. In fact, commercial AM radio signals are only allowed to broadcast a bandwidth of 10 kHz, in order to maximize the number of available stations. (Obviously, two different stations cannot broadcast in frequency ranges which overlap.) This means that commercial AM radio can only carry audible information in the range 0 to about 5 kHz. This is perfectly adequate for ordinary speech, but only barely adequate for music.

Let us now consider how we might transmit a *digital* signal over AM radio. Suppose that each data "bit" in the signal takes the form of a Gaussian envelope, of characteristic duration σ_t, superimposed on a carrier wave whose frequency is ω_0: *i.e.*,

$$\psi(t) = \exp\left(-\frac{t^2}{2\,\sigma_t^2}\right)\,\cos(\omega_0\,t). \tag{8.67}$$

Of course, we must have $\omega_0\,\sigma_t \gg 1$: *i.e.*, the period of the carrier wave must be much less than the duration of the bit. Figure 8.5 illustrates a digital bit calculated for $\omega_0\,\sigma_t = 20$.

The sine Fourier transform of the signal (8.67) is zero by symmetry. However, its cosine Fourier transform takes the form

$$C(k) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2\,\sigma_t^2}\right)\cos(\omega_0\,t)\,\cos(\omega\,t)\,dt, \tag{8.68}$$

$$= \frac{1}{4\pi}\int_{-\infty}^{\infty}\exp\left(-\frac{t^2}{2\,\sigma_t^2}\right)\{\cos[(\omega-\omega_0)\,t]+\cos[(\omega+\omega_0)\,t]\}\,dt.$$

A comparison with Equations (8.15)–(8.18) reveals that

$$C(\omega>0)=\frac{1}{2\,(2\pi\,\sigma_\omega^2)^{1/2}}\,\exp\left(-\frac{(\omega-\omega_0)^2}{2\,\sigma_\omega^2}\right), \tag{8.69}$$

where

$$\sigma_\omega=\frac{1}{\sigma_t}. \tag{8.70}$$

In other words, the Fourier transform of the signal takes the form of a Gaussian in ω-space, which is centered on the carrier frequency, ω_0, and is of characteristic width $\sigma_\omega = 1/\sigma_t$. Thus, the bandwidth of the signal is of order σ_ω. Note that the shorter the signal duration the higher the bandwidth. This is a general rule. A signal of full width at half maximum temporal duration $\Delta t = \sqrt{2\pi}\,\sigma_t$ generally has a Fourier transform of full width at half maximum bandwidth $\Delta\omega = \sqrt{2\pi}\,\sigma_\omega$, so that

$$\Delta\omega\,\Delta t \sim 2\pi. \tag{8.71}$$

This can also be written

$$\Delta f\,\Delta t \sim 1, \tag{8.72}$$

where $\Delta f = \Delta\omega/2\pi$ is the bandwidth in Hertz. The above result is known as the *bandwidth theorem*. Of course, the duration of a digital bit is closely related to the maximum rate with which information can be transmitted in a digital signal. Obviously, the individual bits cannot overlap in time, so the maximum number of bits per second which can be transmitted in a digital signal is of order $1/\Delta t$: *i.e.*, it is of order the bandwidth. Thus, digital signals which transmit information at a rapid rate require large bandwidths: *i.e.*, they occupy a wide range of frequency space.

An old-fashioned black and white TV screen consists of a rectangular grid of black and white spots. A given spot is “white” if the phosphorescent TV screen was recently (*i.e.*, within about $1/50$ th of a second) struck by the electron beam at that location. The spot separation is about 1 mm. A typical screen is $50\,\text{cm} \times 50\,\text{cm}$, and thus has 500 lines with 500 spots per line, or 2.5×10^5 spots. Each spot is renewed every $1/30$ th of a second. (Every other horizontal line is skipped during a given traversal of the electron beam over the screen. The skipped lines are renewed on the next traversal. This technique is known as *interlacing*. Consequently, a given region of the screen, that includes many horizontal lines, has a flicker rate of 60 Hz.) Thus, the rate at which the instructions "turn on" and "turn off" must be sent to the electron beam is $30 \times 2.5 \times 10^5$ or 8×10^6 times a second. The transmitted TV signal must therefore have about 10^7 on-off instruction blips per second. If temporal overlap is to be avoided, each blip can be no longer than $\Delta t \sim 10^{-7}$ seconds in duration. Thus, the required bandwidth is $\Delta f \sim 1/\Delta t \sim 10^7\,\mathrm{Hz} = 10\,\mathrm{MHz}$. The carrier wave frequencies used for conventional broadcast TV lie in the so-called VHF band, and range from about 55 to 210 MHz. Our previous discussion of AM radio might lead us to think that the 10 MHz bandwidth represents the combined extents of an upper and a lower sideband of modulation frequencies. In practice, the carrier wave and one of the sidebands are suppressed: *i.e.*, they are filtered out, and never applied to the antenna. However, they are regenerated in the receiver from the information contained in the single sideband which is broadcast. This technique, which is called *single sideband transmission*, halves the bandwidth requirement to about 5 MHz. Thus, between 55 and 210 MHz there is room for about 30 TV channels, each using a 5 MHz bandwidth. (Actually, there are far fewer TV channels than this in the VHF band, because part of this band is reserved for FM radio, air traffic control, air navigation beacons, marine communications, *etc.*)

8.4 Exercises

1. Verify Equations (8.4)–(8.6). Derive Equations (8.7) and (8.8) from Equation (8.2) and Equations (8.4)–(8.6).

2. Suppose that
$$F(x) = \exp\left(-\frac{x^2}{2\,\sigma_x^2}\right).$$
Demonstrate that
$$\bar{F}(k) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} F(x)\,e^{\,i\,k\,x}\,dx = \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\,\exp\left(-\frac{k^2}{2\,\sigma_k^2}\right),$$
where i is the square-root of minus one, and $\sigma_k = 1/\sigma_x$. [Hint: You will need to complete the square of the exponent of e, transform the variable of integration, and then make use of the standard result that $\int_{-\infty}^{\infty} e^{-y^2}\,dy = \sqrt{\pi}$.] Hence, show from de Moive's theorem, $\exp(\,i\,k\,x) \equiv \cos x + i\,\sin x$, that
$$C(k) \;\equiv\; \frac{1}{2\pi}\int_{-\infty}^{\infty} F(x)\,\cos(k\,x)\,dx = \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\,\exp\left(-\frac{k^2}{2\,\sigma_k^2}\right),$$

$$S(k) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} F(x)\,\sin(k\,x)\,dx = 0.$$

3. Demonstrate that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\,\exp\left(-\frac{k^2}{2\,\sigma_k^2}\right)dk = 1.$$

4. Verify Equations (8.30) and (8.31).

5. Derive Equations (8.11) and (8.12) directly from Equation (8.10) using the results (8.29)–(8.31).

6. Verify directly that (8.45) is a solution of the wave equation (8.32), for arbitrary pulse shapes $F(x)$ and $G(x)$.

7. Verify Equation (8.49).

8. Consider a function $F(t)$ which is zero for negative t, and takes the value $\exp(-t/2\,\tau)$ for $t \geq 0$. Find its Fourier transforms, $C(\omega)$ and $S(\omega)$, defined in

$$F(t) = \int_{-\infty}^{\infty} C(\omega)\,\cos(\omega\,t)\,d\omega + \int_{-\infty}^{\infty} S(\omega)\,\sin(\omega\,t)\,d\omega.$$

9. Suppose that $F(t)$ is zero, except in the interval from $t = -\Delta t/2$ to $t = \Delta t/2$. Suppose that in this interval $F(t)$ makes exactly one sinusoidal oscillation at the angular frequency $\omega_0 = 2\pi/\Delta t$, starting and ending with the value zero. Find the above defined Fourier transforms $C(\omega)$ and $S(\omega)$.

10. Demonstrate that

$$\int_{-\infty}^{\infty} F^2(t)\,dt = 2\pi\int_{-\infty}^{\infty}[C^2(\omega) + S^2(\omega)]\,d\omega,$$

where the relation between $F(t)$, $C(\omega)$, and $S(\omega)$ is defined above. This result is known as *Parseval's theorem*.

11. Suppose that $F(t)$ and $G(t)$ are both even functions of t with the cosine transforms $\bar{F}(\omega)$ and $\bar{G}(\omega)$, so that

$$F(t) = \int_{-\infty}^{\infty} \bar{F}(\omega)\,\cos(\omega\,t)\,d\omega,$$

$$G(t) = \int_{-\infty}^{\infty} \bar{G}(\omega)\,\cos(\omega\,t)\,d\omega.$$

Let $H(t) = F(t)\,G(t)$, and let $\bar{H}(\omega)$ be the cosine transform of this even function, so that

$$H(t) = \int_{-\infty}^{\infty} \bar{H}(\omega)\,\cos(\omega\,t)\,d\omega.$$

Demonstrate that

$$\bar{H}(\omega) = \frac{1}{2}\int_{-\infty}^{\infty} \bar{F}(\omega')\left[\bar{G}(\omega'+\omega) + \bar{G}(\omega'-\omega)\right] d\omega'.$$

This result is known as the *convolution theorem*, since the above type of integral is known as a convolution integral. Suppose that $F(t) = \cos(\omega_0\, t)$. Show that

$$\bar{H}(\omega) = \frac{1}{2}\left[\bar{G}(\omega-\omega_0) + \bar{G}(\omega+\omega_0)\right].$$

9 Dispersive Waves

9.1 Pulse Propagation

Consider a one dimensional wave pulse,

$$\psi(x,t) = \int_{-\infty}^{\infty} C(k)\, \cos(k\,x - \omega\,t)\, dk, \tag{9.1}$$

made up of a linear superposition of cosine waves, with a range of different wavenumbers, all traveling in the positive x-direction. The angular frequency, ω, of each of these waves is related to its wavenumber, k, via the so-called *dispersion relation*, which can be written schematically as

$$\omega = \omega(k). \tag{9.2}$$

In general, this relation is derivable from the wave disturbance's equation of motion. Up to now, we have only considered sinusoidal waves which have *linear* dispersion relations of the form

$$\omega = k\,v, \tag{9.3}$$

where v is a constant. The above expression immediately implies that the waves all have the same *phase velocity*,

$$v_p = \frac{\omega}{k} = v. \tag{9.4}$$

Substituting (9.3) into (9.1), we obtain

$$\psi(x,t) = \int_{-\infty}^{\infty} C(k)\, \cos[k\,(x - v\,t)]\, dk, \tag{9.5}$$

which is clearly the equation of a wave pulse that propagates in the positive x-direction, at the fixed speed v, *without changing shape* (see Chapter 8). The above analysis would seem to suggest that arbitrarily shaped wave pulses generally propagate at the same speed as sinusoidal waves, and do so without dispersing or, otherwise, changing shape. In fact, these statements are only true of pulses made up of superpositions of sinusoidal waves with *linear* dispersion relations. There are, however, many types of sinusoidal wave whose dispersion relations are *nonlinear*. For instance, the dispersion relation of sinusoidal electromagnetic waves propagating through an unmagnetized plasma is (see Section 9.2)

$$\omega = \sqrt{k^2\,c^2 + \omega_p^2}, \tag{9.6}$$

where c is the speed of light in vacuum, and ω_p is a constant, known as the *plasma frequency*, which depends on the properties of the plasma [see Equation (9.28)]. Moreover, the dispersion relation of sinusoidal surface waves in deep water is (see Section 9.4)

$$\omega = \sqrt{g\,k + \frac{T}{\rho}\,k^3}, \tag{9.7}$$

where g is the acceleration due to gravity, T the surface tension, and ρ the mass density. Sinusoidal waves which satisfy nonlinear dispersion relations, such as (9.6) and (9.7), are known as *dispersive waves*, as opposed to waves which satisfy linear dispersion relations, such as (9.3), which are known as *non-dispersive* waves. As we saw above, a wave pulse made up of a linear superposition of non-dispersive sinusoidal waves, all traveling in the same direction, propagates at the common phase velocity of these waves, without changing shape. But, how does a wave pulse made up of a linear superposition of dispersive sinusoidal waves evolve in time?

Suppose that

$$C(k) = \frac{1}{\sqrt{2\pi\,\sigma_k^2}}\,\exp\left(-\frac{(k-k_0)^2}{2\,\sigma_k^2}\right): \tag{9.8}$$

i.e., the function $C(k)$ in (9.1) is a *Gaussian*, of characteristic width σ_k, centered on wavenumber $k = k_0$. It follows, from the properties of the Gaussian function, that $C(k)$ is negligible for $|k-k_0| \gtrsim 3\,\sigma_k$. Thus, the only significant contributions to the wave integral

$$\psi(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\,\sigma_k^2}}\,\exp\left(-\frac{(k-k_0)^2}{2\,\sigma_k^2}\right)\cos(k\,x - \omega\,t)\,dk \tag{9.9}$$

come from a small region of k-space centered on $k = k_0$. Let us Taylor expand the dispersion relation, $\omega = \omega(k)$, about $k = k_0$. Neglecting second-order terms in the expansion, we obtain

$$\omega \simeq \omega(k_0) + (k-k_0)\,\frac{d\omega(k_0)}{dk}. \tag{9.10}$$

It follows that

$$k\,x-\omega\,t\simeq k_0\,x-\omega_0\,t+(k-k_0)\,(x-v_g\,t), \tag{9.11}$$

where $\omega_0=\omega(k_0)$, and

$$v_g=\frac{d\omega(k_0)}{dk} \tag{9.12}$$

is a constant with the dimensions of velocity. Now, if σ_k is sufficiently small then the neglect of second-order terms in the expansion (9.11) is a good approximation, and expression (9.9) becomes

$$\begin{aligned}\psi(x,t) \;\simeq\; & \frac{\cos(k_0\,x-\omega_0\,t)}{\sqrt{2\pi\,\sigma_k^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{(k-k_0)^2}{2\,\sigma_k^2}\right)\cos[(k-k_0)\,(x-v_g\,t)]\,dk\\ & -\frac{\sin(k_0\,x-\omega_0\,t)}{\sqrt{2\pi\,\sigma_k^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{(k-k_0)^2}{2\,\sigma_k^2}\right)\sin[(k-k_0)\,(x-v_g\,t)]\,dk,\end{aligned} \tag{9.13}$$

where use has been made of a standard trigonometric identity. The integral involving $\sin[(k-k_0)\,(x-v_g\,t)]$ is zero, by symmetry. Moreover, an examination of Equations (8.15)–(8.18) reveals that

$$\frac{1}{\sqrt{2\pi\,\sigma_k^2}}\int_{-\infty}^{\infty}\exp\left(-\frac{k^2}{2\,\sigma_k^2}\right)\cos(k\,x)\,dk=\exp\left(-\frac{x^2}{2\,\sigma_x^2}\right), \tag{9.14}$$

where $\sigma_x=1/\sigma_k$. Hence, by analogy with the above expression, (9.13) reduces to

$$\psi(x,t)\simeq\exp\left(-\frac{(x-v_g\,t)^2}{2\,\sigma_x^2}\right)\cos(k_0\,x-\omega_0\,t). \tag{9.15}$$

This is clearly the equation of a wave pulse, of wavenumber k_0 and angular frequency ω_0, with a *Gaussian envelope,* of characteristic width σ_x, whose peak (which is located by setting the argument of the exponential to zero) has the equation of motion

$$x=v_g\,t. \tag{9.16}$$

In other words, the pulse peak—and, hence, the pulse itself—propagates at the velocity v_g, which is known as the *group velocity*. Of course, in the case of non-dispersive waves, the group velocity is the same as the phase velocity (since, if $\omega = k\,v$ then $\omega/k = d\omega/dk = v$). However, for the case of dispersive waves, the two velocities are, in general, *different*.

Equation (9.15) indicates that, as the wave pulse propagates, its envelope remains the same shape. Actually, this result is misleading, and is only obtained because of the neglect of second-order terms in the expansion (9.11). If we keep more terms in this expansion then we can show that the wave pulse does actually change shape as it propagates. However, this demonstration is most readily effected by means of the following simple argument. The pulse extends in Fourier space from $k_0 - \Delta k/2$ to $k_0 + \Delta k/2$, where $\Delta k \sim \sigma_k$. Thus, part of the pulse propagates at the velocity $v_g(k_0 - \Delta k/2)$, and part at the velocity $v_g(k_0 + \Delta k/2)$. Consequently, the pulse *spreads out* as it propagates, since some parts of it move faster than others. Roughly speaking, the spatial extent of the pulse in real space grows as

$$\Delta x \sim (\Delta x)_0 + [v_g(k_0 + \Delta k/2) - v_g(k_0 - \Delta k/2)]\,t \sim (\Delta x)_0 + \frac{dv_g(k_0)}{dk}\,\Delta k\,t, \tag{9.17}$$

where $(\Delta x)_0 \sim \sigma_x = \sigma_k^{-1}$ is the extent of the pulse at $t = 0$. Hence, from (9.12),

$$\Delta x \sim (\Delta x)_0 + \frac{d^2\omega(k_0)}{dk^2}\,\frac{t}{(\Delta x)_0}. \tag{9.18}$$

We, thus, conclude that the spatial extent of the pulse grows *linearly* in time, at a rate proportional to the second derivative of the dispersion relation with respect to k (evaluated at the pulse's central wavenumber). This effect is known as *pulse dispersion*. In summary, a wave pulse made up of a linear superposition of dispersive sinusoidal waves, with a range of different wavenumbers, propagates at the *group velocity*, and also *gradually disperses* as time progresses.

9.2 Electromagnetic Wave Propagation in Plasmas

Consider a point particle of mass m and electric charge q interacting with a sinusoidal electromagnetic wave propagating in the z-direction. Provided that the wave amplitude is not sufficiently large to cause the particle to move at relativistic speeds, the electric component of the wave exerts a much greater force on the particle than the magnetic component. (This follows, from standard electrodynamics, because the ratio of the magnetic to the electric force is of order $B_0\,v/E_0$, where E_0 is the amplitude of the wave electric field-strength, $B_0 = E_0/c$ the amplitude of the wave magnetic field-strength, v the particle velocity, and c the velocity of light in vacuum. Hence, the ratio of the forces is approximately v/c.) Suppose that the electric component of the wave oscillates in the x-direction, and takes the form

$$E_x(z,t) = E_0\,\cos(k\,z - \omega\,t), \tag{9.19}$$

where k is the wavenumber, and ω the angular frequency. The equation of motion of the particle is thus

$$m\,\frac{d^2x}{dt^2} = q\,E_x, \tag{9.20}$$

where x measures its wave-induced displacement in the x-direction. The above equation can easily be solved to give

$$x = -\frac{q\,E_0}{m\,\omega^2}\,\cos(k\,z - \omega\,t). \tag{9.21}$$

Thus, the wave causes the particle to execute sympathetic simple harmonic oscillations, in the x-direction, with an amplitude which is directly proportional to its charge, and inversely proportional to its mass.

Suppose that the wave is actually propagating through an unmagnetized electrically neutral *plasma* consisting of free electrons, of mass m_e and charge $-e$, and free ions, of mass m_i and charge $+e$. Since the plasma is assumed to be electrically neutral, each species must have the same equilibrium number density, n_e. Now, given that the electrons are much less massive than the ions (*i.e.*, $m_e \ll m_i$), but have the same charge (modulo a sign), it follows from (9.21) that the wave-induced oscillations of the electrons are of much higher amplitude than those of the ions. In fact, to a first approximation, we can say that the electrons oscillate whilst the ions remain stationary. Assuming that the electrons and ions are evenly distributed throughout the plasma, the wave-induced displacement of an individual electron generates an effective *electric dipole moment* in the x-direction of the form $p_x = -e\,x$ (the other component of the dipole is a stationary ion of charge $+e$ located at $x = 0$). Hence, the x-directed *electric dipole moment per unit volume* is

$$P_x = n_e\,p_x = -n_e\,e\,x. \tag{9.22}$$

Given that all of the electrons oscillate according to Equation (9.21) (with $q = -e$ and $m = m_e$), we obtain

$$P_x = -\frac{n_e\,e^2\,E_0}{m_e\,\omega^2}\,\cos(k\,z - \omega\,t). \tag{9.23}$$

Now, we saw earlier, in Section 7.7, that the z-directed propagation of an electromagnetic wave, polarized in the x-direction (*i.e.*, with its electric component oscillating in the x-direction), through a dielectric medium is governed by

$$\frac{\partial E_x}{\partial t} = -\frac{1}{\epsilon_0}\left(\frac{\partial P_x}{\partial t} + \frac{\partial H_y}{\partial z}\right), \tag{9.24}$$

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0}\,\frac{\partial E_x}{\partial z}. \tag{9.25}$$

Thus, writing E_x in the form (9.19), H_y in the form

$$H_y(z,t) = Z^{-1}\,E_0\,\cos(k\,z - \omega\,t), \tag{9.26}$$

where Z is the effective *impedance* of the plasma, and P_x in the form (9.23), Equations (9.24) and (9.25) can easily be shown to yield the nonlinear dispersion relation (see Exercise 1)

$$\omega^2 = k^2\,c^2 + \omega_p^2, \tag{9.27}$$

where $c = 1/\sqrt{\epsilon_0\,\mu_0}$ is the velocity of light in vacuum, and the so-called *plasma frequency*,

$$\omega_p = \left(\frac{n_e\,e^2}{\epsilon_0\,m_e}\right)^{1/2}, \tag{9.28}$$

is the characteristic frequency of *collective* electron oscillations in the plasma. Equations (9.24) and (9.25) also yield

$$Z = \frac{Z_0}{n}, \tag{9.29}$$

where $Z_0 = \sqrt{\mu_0/\epsilon_0}$ is the impedance of free space, and

$$n = \frac{k\,c}{\omega} = \left(1 - \frac{\omega_p^2}{\omega^2}\right)^{1/2} \tag{9.30}$$

the effective *refractive index* of the plasma. We, thus, conclude that sinusoidal electromagnetic waves propagating through a plasma have a *nonlinear* dispersion relation. Moreover, it is clear that this nonlinearity arises because the effective refractive index of the plasma is *frequency dependent*.

The expression (9.30) for the refractive index of a plasma has some rather unusual properties. For wave frequencies lying above the plasma frequency (*i.e.*, $\omega > \omega_p$), it yields a real refractive index which is *less than unity*. On the other hand, for wave frequencies lying below the plasma frequency (*i.e.*, $\omega < \omega_p$), it yields an *imaginary* refractive index. Neither of these results makes much sense. The former result is problematic because if the refractive index is less than unity then the wave *phase velocity*, $v_p = \omega/k = c/n$, becomes *superluminal* (*i.e.*, $v_p > c$), and superluminal velocities are generally thought to be unphysical. The latter result is problematic because an imaginary refractive index implies an imaginary phase velocity, which seems utterly meaningless. Let us investigate further.

Consider, first of all, the high frequency limit, $\omega > \omega_p$. According to (9.30), a sinusoidal electromagnetic wave of angular frequency $\omega > \omega_p$ propagates through the plasma at the superluminal phase velocity

$$v_p = \frac{\omega}{k} = \frac{c}{n} = \frac{c}{(1 - \omega_p^2/\omega^2)^{1/2}}. \tag{9.31}$$

But, is this really unphysical? As is well-known, Einstein's theory of relativity forbids *information* from traveling faster than the velocity of light in vacuum, since this would violate *causality* (*i.e.*, it would be possible to transform to a valid frame of reference in which an effect occurs prior to its cause.) However, a sinusoidal wave with a unique frequency, and an infinite spatial extent, does not transmit any information. (Recall, for instance, from Section 8.3, that the carrier wave in an AM radio signal transmits no information.) So, at what speed do electromagnetic waves in a plasma transmit information? Well, the most obvious way of using such waves to transmit information would be to send a message via *Morse code*. In other words, we could transmit a message by means of short *wave pulses*, of varying lengths and interpulse spacings, which are made to propagate through the plasma. The pulses in question would definitely transmit information, so the velocity of information propagation must be the same as that of the pulses: *i.e.*, the *group velocity*, $v_g = d\omega/dk$. Differentiating the dispersion relation (9.27) with respect to k, we obtain

$$2\,\omega\,\frac{d\omega}{dk} = 2\,k\,c^2, \tag{9.32}$$

or

$$\frac{\omega}{k}\,\frac{d\omega}{dk} = v_p\,v_g = c^2. \tag{9.33}$$

Thus, it follows, from (9.31), that the group velocity of high frequency electromagnetic waves in a plasma is

$$v_g = n\,c = (1 - \omega_p^2/\omega^2)^{1/2}\,c. \tag{9.34}$$

Note that the group velocity is *subluminal* (*i.e.*, $v_g < c$). Hence, as long as we accept that high frequency electromagnetic waves transmit information through a plasma at the group velocity, rather than the phase velocity, then there is no problem with causality. Incidentally, it should be clear, from this discussion, that the phase velocity of dispersive waves has very little physical significance. It is the group velocity which matters. For instance, according to Equations (7.120), (9.29), (9.30), and (9.34), the mean flux of electromagnetic energy in the z-direction due to a high frequency sinusoidal wave propagating through a plasma is given by

$$\langle \mathcal{I} \rangle = \frac{1}{2}\,\epsilon_0\,E_0^{\,2}\,n\,c = \frac{1}{2}\,\epsilon_0\,E_0^{\,2}\,v_g, \tag{9.35}$$

since $Z_0 = \sqrt{\mu_0/\epsilon_0}$ and $c = 1/\sqrt{\epsilon_0\,\mu_0}$. Thus, if the group velocity is zero, as is the case when $\omega = \omega_p$, then there is zero energy flux associated with the wave.

The fact that the energy flux and the group velocity of a sinusoidal wave propagating through a plasma both go to zero when $\omega = \omega_p$ suggests that the wave ceases to propagate at all in the low frequency limit, $\omega < \omega_p$. This observation leads us to search for spatially decaying standing wave solutions to (9.24) and (9.25) of the form,

$$E_x(z,t) = E_0\,e^{-k\,z}\,\cos(\omega\,t), \tag{9.36}$$

$$H_y(z,t) = Z^{-1}\,E_0\,e^{-k\,z}\,\sin(\omega\,t). \tag{9.37}$$

It follows from (9.20) and (9.22) that

$$P_x = -\frac{n_e\,e^2\,E_0}{m_e\,\omega^2}\,e^{-k\,z}\,\cos(\omega\,t). \tag{9.38}$$

Substitution into Equations (9.24) and (9.25) reveals that (9.36) and (9.37) are indeed the correct solutions when $\omega < \omega_p$ (see Exercise 2), and also yields

$$k\,c = \sqrt{\omega_p^{\,2} - \omega^2}, \tag{9.39}$$

as well as

$$Z = Z_0\,\frac{\omega}{k\,c} = Z_0\,(\omega_p^{\,2}/\omega^2 - 1)^{-1/2}. \tag{9.40}$$

Furthermore, the mean z-directed electromagnetic energy flux becomes

$$\langle \mathcal{I} \rangle = \langle E_x\,H_y \rangle = E_0^{\,2}\,Z^{-1}\,e^{-2\,k\,z}\,\langle \cos(\omega\,t)\,\sin(\omega\,t) \rangle = 0. \tag{9.41}$$

The above analysis demonstrates that a sinusoidal electromagnetic wave cannot propagate through a plasma when its frequency lies below the plasma frequency. Instead, the amplitude of the wave *decays exponentially* into the plasma. Moreover, the electric and magnetic components of the wave oscillate in *phase quadrature* (*i.e.*, $\pi/2$ radians out of phase), and the wave consequentially has *zero* associated net energy flux. This suggests that a plasma *reflects*, rather than absorbs, an incident electromagnetic wave whose frequency is less than the plasma frequency (since if the wave were absorbed then there would be a net flux of energy into the plasma). Let us investigate what happens when a low frequency electromagnetic wave is normally incident on a plasma in more detail.

Suppose that the region $z < 0$ is a vacuum, and the region $z > 0$ is occupied by a plasma of plasma frequency ω_p. Let the wave electric and magnetic fields in the vacuum region take the form

$$E_x(z,t) = E_i \cos[(\omega/c)(z - c\,t)] + E_r \cos[(\omega/c)(z + c\,t) + \phi_r], \tag{9.42}$$

$$H_y(z,t) = E_i Z_0^{-1} \cos[(\omega/c)(z - c\,t)] - E_r Z_0^{-1} \cos[(\omega/c)(z + c\,t) + \phi_r]. \tag{9.43}$$

Here, E_i is the amplitude of an electromagnetic wave of frequency $\omega < \omega_p$ which is incident on the plasma, whereas E_r is the amplitude of the reflected wave, and ϕ_r the phase of this wave with respect to the incident wave. Moreover, we have made use of the vacuum dispersion relation $\omega = k\,c$. The wave electric and magnetic fields in the plasma are written

$$E_x(z,t) = E_t\, e^{-(\omega/c)\,\alpha z} \cos(\omega\,t + \phi_t), \tag{9.44}$$

$$H_y(z,t) = E_t\, Z_0^{-1}\, \alpha\, e^{-(\omega/c)\,\alpha z} \sin(\omega\,t + \phi_t), \tag{9.45}$$

where E_t is the amplitude of the decaying wave which penetrates into the plasma, ϕ_t is the phase of this wave with respect to the incident wave, and

$$\alpha = \left(\frac{\omega_p^2}{\omega^2} - 1\right)^{1/2}. \tag{9.46}$$

The appropriate matching conditions are the continuity of E_x and H_y at $z = 0$: *i.e.*,

$$E_i \cos(\omega\,t) + E_r \cos(\omega\,t + \phi_r) = E_t \cos(\omega\,t + \phi_t), \tag{9.47}$$

$$E_i \cos(\omega\,t) - E_r \cos(\omega\,t + \phi_r) = E_t\, \alpha \sin(\omega\,t + \phi_t). \tag{9.48}$$

These two equations, which must be satisfied at all times, can be solved to give (see Exercise 3)

$$
\begin{aligned}
E_r &= E_i, && (9.49)\\
\tan\phi_r &= \frac{2\,\alpha}{1-\alpha^2}, && (9.50)\\
E_t &= \frac{2\,E_i}{(1+\alpha^2)^{1/2}}, && (9.51)\\
\tan\phi_t &= \alpha. && (9.52)
\end{aligned}
$$

Thus, the coefficient of reflection,

$$R = \left(\frac{E_r}{E_i}\right)^2 = 1, \tag{9.53}$$

is *unity*, which implies that all of the incident wave energy is reflected by the plasma, and there is no energy absorption. The relative phase of the reflected wave varies from 0 (when $\omega = \omega_p$) to π (when $\omega \ll \omega_p$) radians.

The outer regions of the Earth's atmosphere consist of a tenuous gas which is *partially ionized* by ultraviolet and X-ray radiation from the Sun, as well as by cosmic rays incident from outer space. This region, which is known as the *ionosphere*, acts like a plasma as far as its interaction with radio waves is concerned. The ionosphere consists of many layers. The two most important, as far as radio wave propagation is concerned, are the *E layer*, which lies at an altitude of about 90 to 120 km above the Earth's surface, and the *F layer*, which lies at an altitude of about 120 to 400 km. The plasma frequency in the F layer is generally larger than that in the E layer, because of the greater density of free electrons in the former (recall that $\omega_p \propto \sqrt{n_e}$). The free electron number density in the E layer drops steeply after sunset, due to the lack of solar ionization combined with the gradual recombination of free electrons and ions. Consequently, the plasma frequency in the E layer also drops steeply after sunset. Recombination in the F layer occurs at a much slower rate, so there is nothing like as great a reduction in the plasma frequency of this layer at night. Very High Frequency (VHF) radio signals (*i.e.*, signals with frequencies greater than 30 MHz), which include FM radio and TV signals, have frequencies well in excess of the plasma frequencies of both the E and the F layers, and thus pass straight through the ionosphere. Short Wave (SW) radio signals (*i.e.*, signals with frequencies in the range 3 to 30 MHz) have frequencies in excess of the plasma frequency of the E layer, but not of the F layer. Hence, SW signals

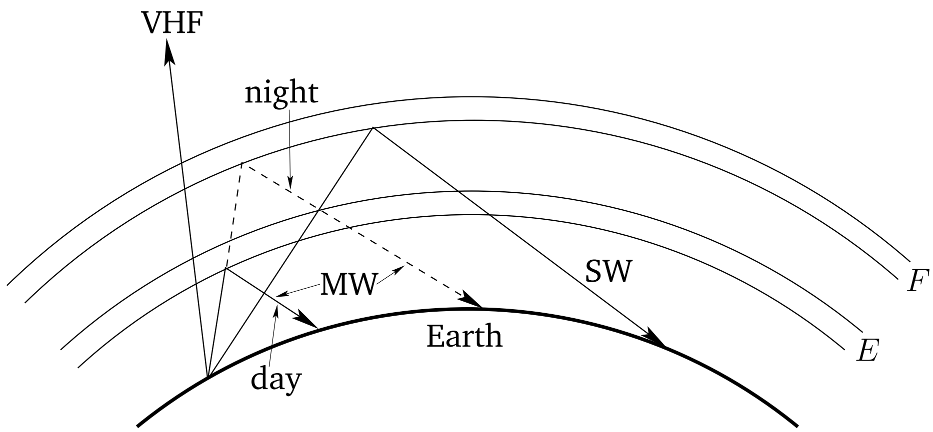


Figure 9.1: *Reflection and transmission of radio waves by the ionosphere.*

pass through the E layer, but are reflected by the F layer. Finally, Medium Wave (MW) radio signals (*i.e.*, signals with frequencies in the range 0.5 to 3 MHz) have frequencies which lie below the plasma frequency of the F layer, and also lie below the plasma frequency of the E layer during daytime, but not during nighttime. Thus, MW signals are reflected by the E layer during the day, but pass through the E layer, and are reflected by the F layer, during the night.

The reflection and transmission of the various different types of radio wave by the ionosphere is shown schematically in Figure 9.1. This diagram explains many of the features of radio reception. For instance, due to the curvature of the Earth's surface, VHF reception is only possible when the receiving antenna is in the line of sight of the transmitting antenna, and is consequently fairly local in nature. MW reception is possible over much larger distances, because the signal is reflected by the ionosphere back towards the Earth's surface. Moreover, long range MW reception improves at night, since the signal is reflected at a higher altitude. Finally, SW radio reception is possible over very large distances, because the signal is reflected at extremely high altitudes.

9.3 Electromagnetic Wave Propagation in Conductors

A so-called *Ohmic conductor* is a medium that satisfies *Ohm's law*, which can be written in the form

$$j = \sigma\, E, \tag{9.54}$$

where j is the current density (*i.e.*, the current per unit area), E the electric field-strength, and σ a constant known as the *conductivity* of the medium in question. (Of course, the current generally flows in the same direction as the electric field.) The z-directed propagation of an electromagnetic wave, polarized in the x-direction, through an Ohmic conductor of conductivity σ is governed by

$$\frac{\partial E_x}{\partial t} + \frac{\sigma}{\epsilon_0}\,E_x = -\frac{1}{\epsilon_0}\,\frac{\partial H_y}{\partial z}, \tag{9.55}$$

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu_0}\,\frac{\partial E_x}{\partial z}, \tag{9.56}$$

For a so-called *good conductor*, which satisfies the inequality $\sigma \gg \epsilon_0\,\omega$, the first term on the left-hand side of Equation (9.55) is negligible with respect to the second term, and the above two equations can be shown to reduce to

$$\frac{\partial E_x}{\partial t} = \frac{1}{\mu_0\,\sigma}\,\frac{\partial^2 E_x}{\partial z^2}, \tag{9.57}$$

$$\frac{\partial H_y}{\partial t} = \frac{1}{\mu_0}\,\frac{\partial E_x}{\partial z}. \tag{9.58}$$

These equations can be solved to give (see Exercise 4)

$$E_x(z,t) = E_0\,e^{-z/d}\,\cos(\omega\,t - z/d), \tag{9.59}$$

$$H_y(z,t) = E_0\,Z^{-1}\,e^{-z/d}\,\cos(\omega\,t - z/d - \pi/4), \tag{9.60}$$

where

$$d = \left(\frac{2}{\mu_0\,\sigma\,\omega}\right)^{1/2}, \tag{9.61}$$

and

$$Z = \left(\frac{\omega\,\mu_0}{\sigma}\right)^{1/2} = \left(\frac{\omega\,\epsilon_0}{\sigma}\right)^{1/2} Z_0. \tag{9.62}$$

Equations (9.59) and (9.60) indicate that the amplitude of an electromagnetic wave propagating through a conductor *decays exponentially* on a characteristic lengthscale, d, which is known as the *skin-depth*. Consequently, an electromagnetic wave cannot penetrate more than a few skin-depths into a conducting medium. Note that the skin-depth is smaller at higher frequencies. This implies that high frequency waves penetrate a shorter distance into a conductor than low frequency waves.

Consider a typical metallic conductor such as copper, whose electrical conductivity at room temperature is about $6 \times 10^7 \, (\Omega \, \mathrm{m})^{-1}$. Copper, therefore, acts as a good conductor for all electromagnetic waves of frequency below about 10^{18} Hz. The skin-depth in copper for such waves is thus

$$d = \sqrt{\frac{2}{\mu_0 \, \sigma \, \omega}} \simeq \frac{6}{\sqrt{f(\mathrm{Hz})}} \, \mathrm{cm}. \tag{9.63}$$

It follows that the skin-depth is about 6 cm at 1 Hz, but only about 2 mm at 1 kHz. This gives rise to the so-called *skin-effect* in copper wires, by which an oscillating electromagnetic signal of increasing frequency, transmitted along such a wire, is confined to an increasingly narrow layer (whose thickness is of order the skin-depth) on the surface of the wire.

The conductivity of sea-water is only about $\sigma \simeq 5 \, (\Omega \, \mathrm{m})^{-1}$. However, this is still sufficiently high for sea-water to act as a good conductor for all radio frequency electromagnetic waves (*i.e.*, $f = \omega/2\pi < 1$ GHz). The skin-depth at 1 MHz ($\lambda \sim 2$ km) is about 0.2 m, whereas that at 1 kHz ($\lambda \sim 2000$ km) is still only about 7 m. This obviously poses quite severe restrictions for radio communication with submerged submarines. Either the submarines have to come quite close to the surface to communicate (which is dangerous), or the communication must be performed with extremely low frequency (ELF) waves (*i.e.*, $f < 100$ Hz). Unfortunately, such waves have very large wavelengths ($\lambda > 20,000$ km), which means that they can only be efficiently generated by gigantic antennas.

According to Equation (9.60), the phase of the magnetic component of an electromagnetic wave propagating through a good conductor lags that of the electric component by $\pi/4$ radians. It follows that the mean energy flux into the conductor takes the form

$$\begin{aligned} \langle \mathcal{I} \rangle &= \langle E_x \, H_y \rangle = |E_x|^2 \, Z^{-1} \, \langle \cos(\omega \, t - z/d) \, \cos(\omega \, t - z/d - \pi/4) \rangle \\ &= \frac{|E_x|^2}{\sqrt{8} \, Z}, \end{aligned} \tag{9.64}$$

where $|E_x| = E_0 \, e^{-z/d}$ is the amplitude of the electric component of the wave. The fact that the mean energy flux is *positive* indicates that part of the wave energy is absorbed by the conductor. In fact, the absorbed energy corresponds to the energy lost due to Ohmic heating in the conductor (see Exercise 5).

Note, from (9.62), that the impedance of a good conductor is far less than that of a vacuum (*i.e.*, $Z \ll Z_0$). This implies that the ratio of the magnetic to the electric components of an electromagnetic wave propagating through a good conductor is far larger than that of a wave propagating through a vacuum.

Suppose that the region $z < 0$ is a vacuum, and the region $z > 0$ is occupied by a good conductor of conductivity σ. Let the wave electric and magnetic fields in the vacuum region take the form of the incident and reflected waves specified in (9.42) and (9.43). The wave electric and magnetic fields in the conductor are written

$$E_x(z,t) = E_t\, e^{-z/d} \cos(\omega\, t - z/d + \phi_t), \tag{9.65}$$

$$H_y(z,t) = E_t\, Z_0^{-1}\, \alpha^{-1}\, e^{-z/d} \cos(\omega\, t - z/d - \pi/4 + \phi_t), \tag{9.66}$$

where E_t is the amplitude of the decaying wave which penetrates into the conductor, ϕ_t is the phase of this wave with respect to the incident wave, and

$$\alpha = \frac{Z}{Z_0} = \left(\frac{\epsilon_0\, \omega}{\sigma}\right)^{1/2} \ll 1. \tag{9.67}$$

The appropriate matching conditions are the continuity of E_x and H_y at $z = 0$: *i.e.*,

$$E_i \cos(\omega\, t) + E_r \cos(\omega\, t + \phi_r) = E_t \cos(\omega\, t + \phi_t), \tag{9.68}$$

$$\alpha\, [E_i \cos(\omega\, t) - E_r \cos(\omega\, t + \phi_r)] = E_t \cos(\omega\, t - \pi/4 + \phi_t). \tag{9.69}$$

Equations (9.68) and (9.69), which must be satisfied at all times, can be solved, in the limit $\alpha \ll 1$, to give (see Exercise 6)

$$E_r \simeq -(1 - \sqrt{2}\, \alpha)\, E_i, \tag{9.70}$$

$$\phi_r \simeq -\sqrt{2}\, \alpha, \tag{9.71}$$

$$E_t \simeq 2\, \alpha\, E_i, \tag{9.72}$$

$$\phi_t \simeq \pi/4. \tag{9.73}$$

Hence, the coefficient of reflection becomes

$$R \simeq \left(\frac{E_r}{E_i}\right)^2 \simeq 1 - 2\sqrt{2}\, \alpha = 1 - \left(\frac{8\, \epsilon_0\, \omega}{\sigma}\right)^{1/2}. \tag{9.74}$$

According to the above analysis, a good conductor reflects a normally incident electromagnetic wave with a phase shift of almost π radians (*i.e.*,

$E_r \simeq -E_i$). The coefficient of reflection is just less than unity, indicating that, whilst most of the incident energy is reflected by the conductor, a small fraction of it is absorbed.

High-quality metallic mirrors are generally coated in silver, whose conductivity is $6.3\times10^7\,(\Omega\,\mathrm{m})^{-1}$. It follows, from (9.74), that at optical frequencies ($\omega = 4\times10^{15}$ rad./s) the coefficient of reflection of a silvered mirror is $R \simeq 93.3\%$. This implies that about 7% of the light incident on the mirror is absorbed, rather than being reflected. This rather severe light loss can be problematic in instruments, such as astronomical telescopes, which are used to view faint objects.

9.4 Surface Wave Propagation in Water

Consider a stationary body of water, of uniform depth d, located on the surface of the Earth. Let us find the dispersion relation of a wave propagating across the water's surface. Suppose that the Cartesian coordinate x measures horizontal distance, whilst the coordinate z measures vertical height, with $z = 0$ corresponding to the unperturbed surface of the water. Let there be no variation in the y-direction: *i.e.*, let the wavefronts of the wave all be parallel to the y-axis. Finally, let $v_x(x, z, t)$ and $v_z(x, z, t)$ be the perturbed horizontal and vertical velocity fields of the water due to the wave. It is assumed that there is no motion in the y-direction.

Now, water is essentially *incompressible*. Thus, any wave disturbance in water is constrained to preserve the volume of a co-moving volume element. Equivalently, the inflow rate of water into a stationary volume element must match the outflow rate. Consider a stationary cubic volume element lying between x and $x + dx$, y and $y + dy$, and z and $z + dz$. The element has two faces, of area $dy\,dz$, perpendicular to the x-axis, located at x and $x + dx$. Water flows into the element through the former face at the rate $v_x(x, z, t)\,dy\,dz$ (*i.e.*, the product of the area of the face and the normal velocity), and out of the element though the latter face at the rate $v_x(x + dx, z, t)\,dy\,dz$. The element also has two faces perpendicular to the y-axis, but there is no flow through these faces, since $v_y = 0$. Finally, the element has two faces, of area $dx\,dy$, perpendicular to the z-axis, located at z and $z + dz$. Water flows into the element through the former face at the rate $v_z(x, z, t)\,dx\,dy$, and out of the element though the latter face at the rate $v_z(x, z + dz, t)\,dx\,dy$. Thus, the net rate at which water flows into the element is $v_x(x, z, t)\,dy\,dz + v_z(x, z, t)\,dx\,dy$. Likewise, the net rate at which water flows out of the element is $v_x(x+dx, z, t)\,dy\,dz + v_z(x, z+dz, t)\,dx\,dy$.

If the water is to remain incompressible then the inflow and outflow rates must match: *i.e.*,

$$v_x(x,z,t)\,dy\,dz+v_z(x,z,t)\,dx\,dy = v_x(x+dx,z,t)\,dy\,dz+v_z(x,z+dz,t)\,dx\,dy, \tag{9.75}$$

or

$$\left(\frac{v_x(x+dx,z,t)-v_x(x,z,t)}{dx}+\frac{v_z(x,z+dz,t)-v_z(x,z,t)}{dz}\right)dx\,dy\,dz=0. \tag{9.76}$$

Hence, the incompressibility constraint reduces to

$$\frac{\partial v_x}{\partial x}+\frac{\partial v_z}{\partial z}=0. \tag{9.77}$$

Consider the equation of motion of a small volume element of water lying between x and $x+dx$, y and $y+dy$, and z and $z+dz$. The mass of this element is $\rho\,dx\,dy\,dz$, where ρ is the uniform mass density of water. Suppose that $p(x,z,t)$ is the *pressure* in the water, which is assumed to be *isotropic*. The net horizontal force on the element is $p(x,z,t)\,dy\,dz-p(x+dx,z,t)\,dy\,dz$ (since force is pressure times area, and the external pressure forces acting on the element act inward across its surface). Hence, the element's horizontal equation of motion is

$$\rho\,dx\,dy\,dz\,\frac{\partial v_x(x,z,t)}{\partial t}=-\left(\frac{p(x+dx,z,t)-p(x,z,t)}{dx}\right)dx\,dy\,dz, \tag{9.78}$$

which reduces to

$$\rho\,\frac{\partial v_x}{\partial t}=-\frac{\partial p}{\partial x}. \tag{9.79}$$

The vertical equation of motion is similar, except that the element is subject to a downward acceleration, g, due to gravity. Hence, we obtain

$$\rho\,\frac{\partial v_z}{\partial t}=-\frac{\partial p}{\partial z}-\rho\,g. \tag{9.80}$$

Now, we can write

$$p=p_0-\rho\,g\,z+p_1, \tag{9.81}$$

where p_0 is atmospheric pressure (*i.e.*, the pressure at the surface of the water), and p_1 is the pressure perturbation due to the wave. Of course, in the absence of the wave, the water pressure at a depth h below the surface is $p_0+\rho\,g\,h$. Substitution into (9.79) and (9.80) yields

$$\rho\,\frac{\partial v_x}{\partial t} = -\frac{\partial p_1}{\partial x}, \tag{9.82}$$

$$\rho\,\frac{\partial v_z}{\partial t} = -\frac{\partial p_1}{\partial z}. \tag{9.83}$$

It follows that

$$\rho\,\frac{\partial^2 v_x}{\partial z\,\partial t} - \rho\,\frac{\partial^2 v_z}{\partial x\,\partial t} = -\frac{\partial^2 p_1}{\partial z\,\partial x} + \frac{\partial^2 p_1}{\partial x\,\partial z} = 0, \tag{9.84}$$

which implies that

$$\rho\,\frac{\partial}{\partial t}\left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) = 0, \tag{9.85}$$

or

$$\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x} = 0. \tag{9.86}$$

(Actually, the above quantity could be non-zero and constant in time, but this is not consistent with an oscillating wave-like solution.)

Equation (9.86) is automatically satisfied if

$$v_x = \frac{\partial \phi}{\partial x}, \tag{9.87}$$

$$v_z = \frac{\partial \phi}{\partial z}. \tag{9.88}$$

Equation (9.77) then gives

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2} = 0. \tag{9.89}$$

Finally, Equations (9.82) and (9.83) yield

$$p_1 = -\rho\,\frac{\partial \phi}{\partial t}. \tag{9.90}$$

As we have just seen, surface waves in water are governed by Equation (9.89), which is known as *Laplace's equation*. We next need to derive the physical constraints which must be satisfied by the solution to this equation at the water's upper and lower boundaries. Now, the water is bounded from below by a solid surface located at $z = -d$. Assuming that the water always remains in contact with this surface, the appropriate physical constraint at the lower boundary is $v_z(x, -d, t) = 0$ (*i.e.*, there is no vertical motion of the water at the lower boundary), or

$$\left.\frac{\partial \phi}{\partial z}\right|_{z=-d} = 0. \tag{9.91}$$

The physical constraint at the water's upper boundary is a little more complicated, since this boundary is a free surface. Let $\zeta(x, t)$ represent the vertical displacement of the water's surface. It follows that

$$\frac{\partial \zeta}{\partial t} = v_z|_{z=0} = \left.\frac{\partial \phi}{\partial z}\right|_{z=0}. \tag{9.92}$$

Now, the physical constraint at the surface is that the water pressure be equal to atmospheric pressure, since there cannot be a pressure discontinuity across a free surface. Thus, it follows from (9.81) that

$$p_0 = p_0 - \rho\, g\, \zeta(x,t) + p_1(x,0,t). \tag{9.93}$$

Finally, differentiating with respect to t, and making use of Equations (9.90) and (9.92), we obtain

$$\left.\frac{\partial\phi}{\partial z}\right|_{z=0} = -g^{-1}\left.\frac{\partial^2\phi}{\partial t^2}\right|_{z=0}. \tag{9.94}$$

Hence, the problem boils down to solving Laplace's equation, (9.89), subject to the physical constraints (9.91) and (9.94).

Let us search for a propagating wave-like solution of (9.89) of the form

$$\phi(x,z,t) = F(z)\,\cos(k\,x - \omega\,t). \tag{9.95}$$

Substitution into (9.89) yields

$$\frac{d^2F}{dz^2} - k^2\,F = 0, \tag{9.96}$$

whose independent solutions are $\exp(+k\,z)$ and $\exp(-k\,z)$. Hence, the most general wavelike solution to Laplace's equation takes the form

$$\phi(x,z,t) = A\,e^{k\,z}\cos(k\,x-\omega\,t) + B\,e^{-k\,z}\cos(k\,x-\omega\,t), \tag{9.97}$$

where A and B are arbitrary constants. The boundary condition (9.91) is satisfied provided that $B = A\,\exp(-2\,k\,d)$, giving

$$\phi(x,z,t) = A\left[e^{k\,z} + e^{-k\,(z+2d)}\right]\cos(k\,x-\omega\,t), \tag{9.98}$$

The boundary condition (9.94) yields

$$A\,k\left[1-e^{-2\,k\,d}\right]\cos(k\,x-\omega\,t) = A\,\frac{\omega^2}{g}\left[1+e^{-2\,k\,d}\right]\cos(k\,x-\omega\,t), \tag{9.99}$$

which reduces to the dispersion relation

$$\omega^2 = g\,k\,\tanh(k\,d), \tag{9.100}$$

where the function

$$\tanh x \equiv \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{9.101}$$

is known as a *hyperbolic tangent*.

In *shallow water* (*i.e.*, $k\,d \ll 1$), Equation (9.100) reduces to the linear dispersion relation

$$\omega = k\,\sqrt{g\,d}. \tag{9.102}$$

Here, use has been made of the small argument expansion $\tanh x \simeq x$ for $|x| \ll 1$. We, thus, conclude that surface waves in shallow water are *non-dispersive* in nature, and propagate at the phase velocity $\sqrt{g\,d}$. On the other hand, in *deep water* (*i.e.*, $k\,d \gg 1$), Equation (9.100) reduces to the nonlinear dispersion relation

$$\omega = \sqrt{k\,g}. \tag{9.103}$$

Here, use has been made of the large argument expansion $\tanh x \simeq 1$ for $x \gg 1$. Hence, we conclude that surface waves in deep water are *dispersive* in nature. The phase velocity of the waves is $v_p = \omega/k = \sqrt{g/k}$, whereas the group velocity is $v_g = d\omega/dk = (1/2)\,\sqrt{g/k} = v_p/2$. In other words, the group velocity is *half* the phase velocity, and is largest for long wavelength (*i.e.*, small k) waves.

Water in contact with air actually possesses a *surface tension* $T \simeq 7 \times 10^{-2}\,\mathrm{N\,m^{-1}}$ which allows there to be a small pressure discontinuity across a free surface that is curved. In fact,

$$[p]_{z=0_-}^{z=0_+} = -T\,\frac{\partial^2\zeta}{\partial x^2}. \tag{9.104}$$

Here, $(\partial\zeta/\partial x^2)^{-1}$ is the *radius of curvature* of the surface. Thus, in the presence of surface tension, the boundary condition (9.93) takes the modified form

$$-\,T\,\frac{\partial^2\zeta}{\partial x^2} = -\rho\,g\,\zeta + p_1|_{z=0}\,, \tag{9.105}$$

which reduces to

$$\left.\frac{\partial\phi}{\partial z}\right|_{z=0} = \frac{T}{\rho\,g}\,\left.\frac{\partial^3\phi}{\partial x^2\,\partial z}\right|_{z=0} - \frac{1}{g}\,\left.\frac{\partial^2\phi}{\partial t^2}\right|_{z=0}. \tag{9.106}$$

This boundary condition can be combined with the solution (9.98), in the deep water limit $k\,d \gg 1$, to give the modified deep water dispersion relation

$$\omega = \sqrt{g\,k + \frac{T}{\rho}\,k^3}. \tag{9.107}$$

Hence, the phase velocity of the waves takes the form

$$v_p = \frac{\omega}{k} = \sqrt{\frac{g}{k} + \frac{T}{\rho}\,k}, \tag{9.108}$$

and the ratio of the group velocity to the phase velocity can be shown to be

$$\frac{v_g}{v_p} = \frac{1}{2}\left[\frac{1 + 3\,T\,k^2/(\rho\,g)}{1 + T\,k^2/(\rho\,g)}\right]. \tag{9.109}$$

Thus, the phase velocity attains a minimum value of $\sqrt{2}\,(g\,T/\rho)^{1/4} \sim 0.2\,\mathrm{m\,s^{-1}}$ when $k = k_0 \equiv (\rho\,g/T)^{1/2}$, which corresponds to $\lambda \sim 2\,\mathrm{cm}$. The group velocity equals the phase velocity at this wavelength. For long wavelength waves (*i.e.*, $k \ll k_0$), gravity dominates surface tension, the phase velocity scales as $k^{-1/2}$, and the group velocity is half the phase velocity. On the other hand, for short wavelength waves (*i.e.*, $k \gg k_0$), surface tension dominates gravity, the phase velocity scales as $k^{1/2}$, and the group velocity is $3/2$ times the phase velocity. The fact that the phase velocity and the group velocity both attain minimum values when $\lambda \sim 2\,\mathrm{cm}$ means that when a wave disturbance containing a wide spectrum of wavelengths, such as might be generated by throwing a rock into the water, travels across the surface of a lake, and reaches the shore, the short and long wavelength components of the disturbance generally arrive before the components of intermediate wavelength.

9.5 Exercises

1. Derive expressions (9.27) and (9.29) for propagating electromagnetic waves in a plasma from Equations (9.20), (9.22), (9.24), and (9.25).

2. Derive expressions (9.39) and (9.40) for evanescent electromagnetic waves in a plasma from Equations (9.20), (9.22), (9.24), and (9.25).

3. Derive Equations (9.49)–(9.52) from Equations (9.47) and (9.48).

4. Derive Equations (9.59)–(9.62) from Equations (9.57) and (9.58).

5. Consider an electromagnetic wave propagating in the positive z-direction through a conducting medium of conductivity σ. Suppose that the wave electric field is
$$E_x(z,t) = E_0\,e^{-z/d}\,\cos(\omega\,t - z/d),$$
where d is the skin-depth. Demonstrate that the mean electromagnetic energy flux across the plane $z = 0$ matches the mean rate at which electromagnetic energy is dissipated, per unit area, due to Ohmic heating in the region $z > 0$. (The rate of ohmic heating per unit volume is $\sigma\,E_x^{\,2}$).

6. Derive Equations (9.70)–(9.73) from Equations (9.68) and (9.69), in the limit $\alpha \ll 1$.

7. Demonstrate that the phase velocity of traveling waves on an infinitely long beaded string is
$$v_p = v_0 \frac{\sin(k\,a/2)}{(k\,a/2)},$$
where $v_0 = \sqrt{T\,a/m}$, T is the tension in the string, a the spacing between the beads, m the mass of the beads, and k the wavenumber of the wave. What is the group velocity?

8. The number density of free electrons in the ionosphere, n_e, as a function of vertical height, z, is measured by timing how long it takes a radio pulse launched vertically upward from the ground ($z = 0$) to return to ground level again, after reflection by the ionosphere, as a function of the pulse frequency, ω. It is conventional to define the *equivalent height*, $h(\omega)$, of the reflection layer as the height it would need to have off the ground if the pulse always traveled at the velocity of light in vacuum. Demonstrate that
$$h(\omega) = \int_0^{z_0} \frac{dz}{[1 - \omega_p^2(z)/\omega^2]^{1/2}},$$
where $\omega_p^2(z) = n_e(z)\,e^2/(\epsilon_0\,m_e)$, and $\omega_p^2(z_0) = \omega^2$. Show that if $n_e \propto z^p$ then $h \propto \omega^{2/p}$.

9. A uniform rope of mass per unit length ρ and length L hangs vertically. Determine the tension T in the rope as a function of height from the bottom of the rope. Show that the time required for a transverse wave pulse to travel from the bottom to the top of the rope is $2\sqrt{L/g}$.

10. The aluminium foil used in cooking has an electrical conductivity $\sigma = 3.5 \times 10^7\ (\Omega\,\text{m})^{-1}$, and a typical thickness $\delta = 2 \times 10^{-4}$ m. Show that such foil can be used to shield a region from electromagnetic waves of a given frequency, provided that the skin-depth of the waves in the foil is less than about a third of its thickness. Since skin-depth increases as frequency decreases, it follows that the foil can only shield waves whose frequency exceeds a critical value. Estimate this critical frequency (in Hertz). What is the corresponding wavelength?

11. A sinusoidal surface wave travels from deep water toward the shore. Does its wavelength increase, decrease, or stay the same, as it approaches the shore? Explain.

12. Demonstrate that the dispersion relation (9.107) for surface water waves generalizes to
$$\omega^2 = \left(g\,k + \frac{T}{\rho}\,k^3\right)\tanh(k\,d)$$
in water of arbitrary depth.

13. Demonstrate that a small amplitude surface wave, of angular frequency ω and wavenumber k, traveling over the surface of a lake of uniform depth d causes an individual water volume element located at a depth h below the surface to execute a non-propagating elliptical orbit whose major and minor axes are horizontal and vertical, respectively. Show that the variation of the major and minor radii of the orbit with depth is $A\,\cosh[k\,(d-h)]$ and $A\,\sinh[k\,(d-h)]$, respectively, where A is a constant. Demonstrate that the volume elements are moving horizontally in the same direction as the wave at the top of their orbits, and in the opposite direction at the bottom. Show that a surface wave traveling over the surface of a very deep lake causes water volume elements to execute non-propagating circular orbits whose radii decrease exponentially with depth.

10 Multi-Dimensional Waves

10.1 Plane Waves

As we have already seen, a sinusoidal wave of amplitude $\psi_0 > 0$, wavenumber $k > 0$, and angular frequency $\omega > 0$, propagating in the positive x-direction, can be represented by a *wavefunction* of the form

$$\psi(x, t) = \psi_0 \cos(k\,x - \omega\,t). \tag{10.1}$$

Now, the above type of wave is conventionally termed a *one-dimensional plane wave*. It is *one-dimensional* because its associated wavefunction only depends on a single Cartesian coordinate. Furthermore, it is a *plane wave* because the wave maxima, which are located at

$$k\,x - \omega\,t = j\,2\pi, \tag{10.2}$$

where j is an integer, consist of a series of *parallel planes*, normal to the x-axis, which are equally spaced a distance $\lambda = 2\pi/k$ apart, and propagate along the x-axis at the fixed speed $v = \omega/k$. These conclusions follow because Equation (10.2) can be re-written in the form

$$x = d, \tag{10.3}$$

where $d = j\,\lambda + v\,t$. Moreover, (10.3) is clearly the equation of a plane, normal to the x-axis, whose distance of closest approach to the origin is d.

The previous equation can also be written in the coordinate-free form

$$\mathbf{n}\cdot\mathbf{r} = d, \tag{10.4}$$

where $\mathbf{n} = (1, 0, 0)$ is a unit vector directed along the x-axis, and $\mathbf{r} = (x, y, z)$ represents the vector displacement of a general point from the origin. Since there is nothing special about the x-direction, it follows that if $\mathbf{n}$ is re-interpreted as a unit vector pointing in an *arbitrary* direction then (10.4) can be re-interpreted as the general equation of a plane. As before, the plane is normal to $\mathbf{n}$, and its distance of closest approach to the origin is d. See Figure 10.1. This observation allows us to write the three-dimensional equivalent to the wavefunction (10.1) as

$$\psi(x, y, z, t) = \psi_0 \cos(\mathbf{k}\cdot\mathbf{r} - \omega\,t), \tag{10.5}$$

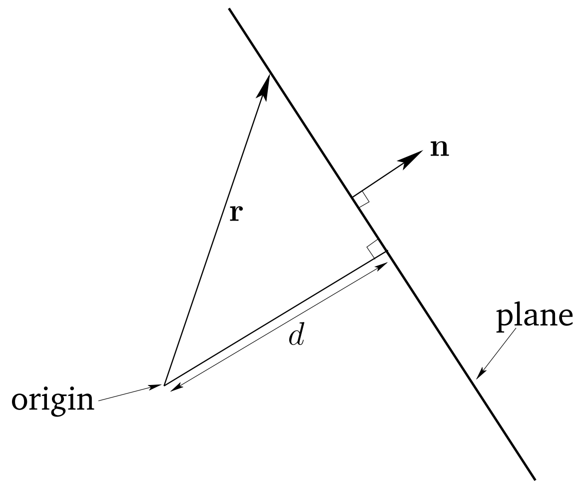


Figure 10.1: *The solution of* $\mathbf{n}\cdot\mathbf{r} = d$ *is a plane.*

where the constant vector $\mathbf{k} = (k_x,\, k_y,\, k_z) = k\,\mathbf{n}$ is called the *wavevector*. The wave represented above is conventionally termed a *three-dimensional plane wave*. It is three-dimensional because its wavefunction, $\psi(x, y, z, t)$, depends on all three Cartesian coordinates. Moreover, it is a plane wave because the wave maxima are located at

$$\mathbf{k}\cdot\mathbf{r} - \omega\, t = j\, 2\pi, \tag{10.6}$$

or

$$\mathbf{n}\cdot\mathbf{r} = j\,\lambda + v\, t, \tag{10.7}$$

where $\lambda = 2\pi/k$, and $v = \omega/k$. Note that the wavenumber, k, is the *magnitude* of the wavevector, $\mathbf{k}$: *i.e.*, $k \equiv |\mathbf{k}|$. It follows, by comparison with Equation (10.4), that the wave maxima consist of a series of parallel planes, normal to the wavevector, which are equally spaced a distance λ apart, and propagate in the $\mathbf{k}$-direction at the fixed speed v. See Figure 10.2. Hence, the direction of the wavevector specifies the wave propagation direction, whereas its magnitude determines the wavenumber, k, and, thus, the wavelength, $\lambda = 2\pi/k$. Actually, the most general expression for the wavefunction of a plane wave is $\psi = \psi_0\cos(\phi + \mathbf{k}\cdot\mathbf{r} - \omega\, t)$, where ϕ is a constant *phase angle*. As is easily appreciated, the inclusion of a non-zero phase angle in the wavefunction merely shifts all the wave maxima a distance $-(\phi/2\pi)\,\lambda$ in the $\mathbf{k}$-direction. In the following, whenever possible, ϕ is set to zero, for the sake of simplicity.

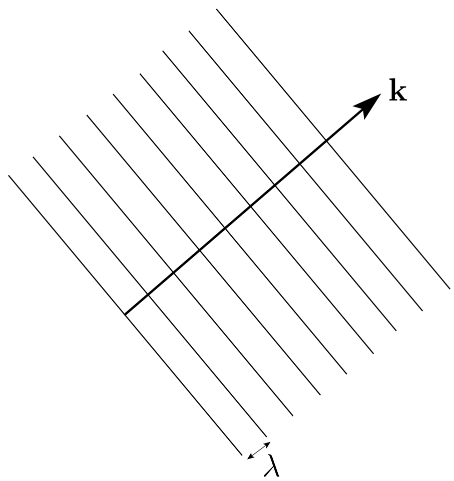


Figure 10.2: *Wave maxima associated with a plane wave.*

10.2 Three-Dimensional Wave Equation

As is readily demonstrated (see Exercise 1), the one-dimensional plane wave solution (10.1) satisfies the *one-dimensional wave equation*,

$$\frac{\partial^2\psi}{\partial t^2} = v^2\,\frac{\partial^2\psi}{\partial x^2}. \tag{10.8}$$

Likewise, the three-dimensional plane wave solution (10.5) satisfies the *three-dimensional wave equation* (see Exercise 1),

$$\frac{\partial^2\psi}{\partial t^2} = v^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi. \tag{10.9}$$

Note that both of these equations are *linear*, and, thus, have *superposable solutions*.

10.3 Laws of Geometric Optics

Suppose that the region $z < 0$ is occupied by a transparent dielectric medium of refractive index n_1, whereas the region $z > 0$ is occupied by a second transparent dielectric medium of refractive index n_2. Let a plane light wave be launched, toward positive z, from a light source of angular frequency ω located at large negative z. Suppose that this so-called *incident* wave has a wavevector $\mathbf{k}_i$. Now, we would expect the incident wave to be *partially reflected* and *partially transmitted* at the interface between the two dielectric media. Let the reflected and transmitted waves have the wavevectors $\mathbf{k}_r$ and $\mathbf{k}_t$, respectively. See Figure 10.3. Hence, we can write

$$\psi(x, y, z, t) = \psi_i \cos(\mathbf{k}_i \cdot \mathbf{r} - \omega\, t) + \psi_r \cos(\mathbf{k}_r \cdot \mathbf{r} - \omega\, t) \tag{10.10}$$

in the region $z < 0$, and

$$\psi(x, y, z, t) = \psi_t \cos(\mathbf{k}_t \cdot \mathbf{r} - \omega\, t) \tag{10.11}$$

in the region $z > 0$. Here, $\psi(x, y, z, t)$ represents the *magnetic* component of the resultant light wave, ψ_i the amplitude of the incident wave, ψ_r the amplitude of the reflected wave, and ψ_t the amplitude of the transmitted wave. Note that all of the component waves have the same angular frequency, ω, since this property is ultimately determined by the wave source. Note, further, that, according to standard electromagnetic theory, if the magnetic component of an electromagnetic wave is specified then the electric component of the wave is fully determined, and can easily be calculated, and *vice versa*.

In general, the wavefunction, ψ, must be *continuous* at $z = 0$, since, according to standard electromagnetic theory, there cannot be a discontinuity in either the normal or the tangential component of a magnetic field across an interface between two (non-magnetic) dielectric media. (Incidentally, the same is not true of an electric field, which can have a normal discontinuity across an interface between two dielectric media. This explains why we have chosen ψ to represent the magnetic, rather than the electric, component of the resultant light wave.) Thus, the matching condition at $z = 0$ takes the form

$$\begin{aligned} &\psi_i \cos(k_{ix}\, x + k_{iy}\, y - \omega\, t) \\ &+\psi_r \cos(k_{rx}\, x + k_{ry}\, y - \omega\, t) = \psi_t \cos(k_{tx}\, x + k_{ty}\, y - \omega\, t). \end{aligned} \tag{10.12}$$

Moreover, this condition must be satisfied at *all* values of x, y, and t. Clearly, this is only possible if

$$k_{ix} = k_{rx} = k_{tx}, \tag{10.13}$$

and

$$k_{iy} = k_{ry} = k_{ty}. \tag{10.14}$$

Suppose that the direction of propagation of the incident wave lies in the x-z plane, so that $k_{iy} = 0$. It immediately follows, from (10.14), that

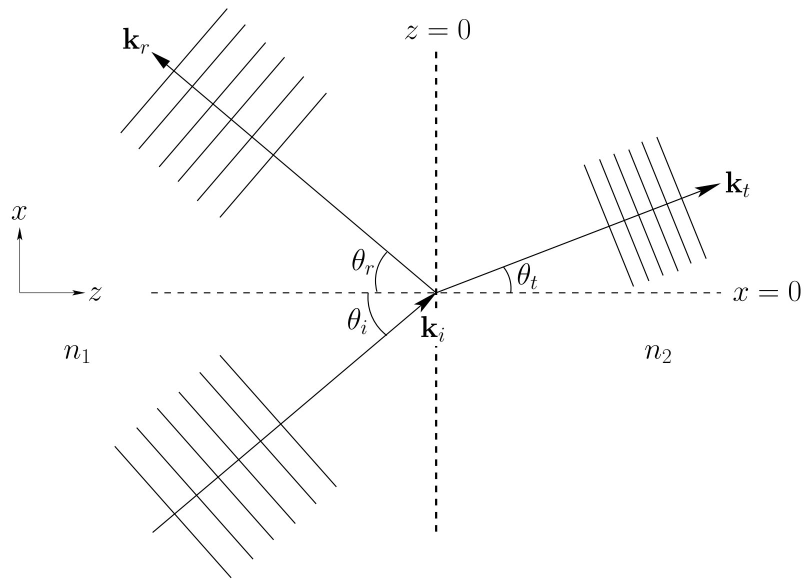


Figure 10.3: *Reflection and refraction of a plane wave at a plane boundary.*

$k_{ry} = k_{ty} = 0$. In other words, the directions of propagation of the reflected and the transmitted waves also lie in the x-z plane. This means that $\mathbf{k}_i$, $\mathbf{k}_r$ and $\mathbf{k}_t$ are *co-planar* vectors. Note that this constraint is implicit in the well-known laws of geometric optics.

Assuming that the above mentioned constraint is satisfied, let the incident, reflected, and transmitted waves subtend angles θ_i, θ_r, and θ_t with the z-axis, respectively. See Figure 10.3. It follows that

$$\mathbf{k}_i = n_1\,k_0\,(\sin\theta_i,\ 0,\ \cos\theta_i), \tag{10.15}$$

$$\mathbf{k}_r = n_1\,k_0\,(\sin\theta_r,\ 0, -\cos\theta_r), \tag{10.16}$$

$$\mathbf{k}_t = n_2\,k_0\,(\sin\theta_t,\ 0,\ \cos\theta_t), \tag{10.17}$$

where $k_0 = \omega/c$ is the vacuum wavenumber, and c the velocity of light in vacuum. Here, we have made use of the fact that wavenumber (*i.e.*, the magnitude of the wavevector) of a light wave propagating through a dielectric medium of refractive index n is $n\,k_0$.

Now, according to Equation (10.13), $k_{ix} = k_{rx}$, which yields

$$\sin\theta_i = \sin\theta_r, \tag{10.18}$$

and $k_{ix} = k_{tx}$, which reduces to

$$n_1\,\sin\theta_i = n_2\,\sin\theta_t. \tag{10.19}$$

The first of these relations states that the angle of incidence, θ_i, is equal to the angle of reflection, θ_r. Of course, this is the familiar *law of reflection*. Moreover, the second relation corresponds to the equally familiar *law of refraction*, otherwise known as *Snell's law*.

Incidentally, the fact that a plane wave propagates through a uniform medium with a *constant* wavevector, and, therefore, a *constant* direction of propagation, is equivalent to the well known *law of rectilinear propagation*, which states that light propagates through a uniform medium in a *straight-line*.

It is clear, from the above discussion, that the laws of geometric optics (*i.e.*, the law of rectilinear propagation, the law of reflection, and the law of refraction) are fully consistent with the wave properties of light, despite the fact that they do not appear to explicitly depend on these properties.

10.4 Waveguides

As we saw in Section 7.5, transmission lines (*e.g.*, ethernet cables) are used to carry high frequency electromagnetic signals over distances which are long compared to the signal wavelength, $\lambda = c/f$, where c is the velocity of light and f the signal frequency (in Hertz). Unfortunately, conventional transmission lines are subject to *radiative losses* (since the lines effectively act as antennas) which increase as the *fourth power* of the signal frequency. Above a certain critical frequency, which typically lies in the microwave band, the radiative losses become intolerably large. Under these circumstances, the transmission line must be replaced by a device known as a *waveguide*. A *waveguide* is basically a long hollow metal box within which electromagnetic signals propagate. Moreover, if the walls of the box are much thicker than the skin-depth (see Section 9.3) in the wall material then the signal is essentially isolated from the outside world, and the radiative losses are consequently negligible.

Consider an evacuated waveguide of rectangular cross-section which runs along the z-direction, and is enclosed by perfectly conducting (*i.e.*, infinite conductivity) metal walls located at $x = 0$, $x = a$, $y = 0$, and $y = b$. Suppose that an electromagnetic wave propagates along the waveguide in the z-direction. For the sake of simplicity, let there be no y-variation of the wave electric or magnetic fields. Now, the wave propagation inside the waveguide is governed by the two-dimensional wave equation [*cf.*, Equation (10.9)]

$$\frac{\partial^2\psi}{\partial t^2} = c^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)\psi, \tag{10.20}$$

where $\psi(x, z, t)$ represents the *electric* component of the wave, which is assumed to be everywhere parallel to the y-axis, and c is the velocity of light in vaccum. The appropriate boundary conditions are

$$\psi(0, z, t) = 0, \tag{10.21}$$

$$\psi(a, z, t) = 0, \tag{10.22}$$

since the electric field inside a perfect conductor is zero (otherwise, an infinite current would flow), and, according to standard electromagnetic theory, there cannot be a tangential discontinuity in the electric field at a conductor/vacuum boundary. (There can, however, be a normal discontinuity. This allows ψ to be non-zero at $y = 0$ and $y = b$.)

Let us search for a separable solution of (10.20) of the form

$$\psi(x, z, t) = \psi_0 \sin(k_x\, x)\, \cos(k\, z - \omega\, t), \tag{10.23}$$

where k represents the z-component of the wavevector (rather than its magnitude), and is the effective wavenumber for propagation along the waveguide. The above solution automatically satisfies the boundary condition (10.21). The second boundary condition (10.22) is satisfied provided

$$k_x = j\,\frac{\pi}{a}, \tag{10.24}$$

where j is a positive integer. Suppose that j takes its smallest possible value 1. (Of course, j cannot be zero, since, in this case, $\psi = 0$ everywhere.) Substitution of expression (10.23) into the wave equation (10.20) yields the dispersion relation

$$\omega^2 = k^2\, c^2 + \omega_0^2, \tag{10.25}$$

where

$$\omega_0 = \frac{\pi\, c}{a}. \tag{10.26}$$

Note that this dispersion relation is analogous in form to the dispersion relation (9.27) for an electromagnetic wave propagating through a plasma, with the *cut-off frequency*, ω_0, playing the role of the plasma frequency, ω_p. The cut-off frequency is so-called because for $\omega < \omega_0$ the wavenumber is imaginary (*i.e.*, $k^2 < 0$), which implies that the wave does not propagate along the waveguide, but, instead, decays exponentially with increasing z.

On the other hand, for wave frequencies above the cut-off frequency the phase velocity,

$$v_p = \frac{\omega}{k} = \frac{c}{\sqrt{1 - \omega_0^2/\omega^2}}, \tag{10.27}$$

is superluminal. This is not a problem, however, since the group velocity,

$$v_g = \frac{d\omega}{dk} = c\sqrt{1 - \omega_0^2/\omega^2}, \tag{10.28}$$

which is the true signal velocity, remains subluminal. (Recall, from Section 9.2, that a high frequency electromagnetic wave propagating through a plasma exhibits similar behavior.) Not surprisingly, the signal velocity goes to zero as $\omega \to \omega_0$, since the signal ceases to propagate at all when $\omega = \omega_0$.

It turns out that waveguides support many distinct modes of propagation. The type of mode discussed above is termed a TE (for transverse electric-field) mode, since the electric field is transverse to the direction of propagation. There are many different sorts of TE mode, corresponding, for instance, to different choices of the mode number, j. However, the $j = 1$ mode has the lowest cut-off frequency. There are also TM (for transverse magnetic-field) modes, and TEM (for transverse electric- and magnetic-field) modes. TM modes also only propagate when the wave frequency exceeds a cut-off frequency. On the other hand, TEM modes (which are the same type of mode as that supported by a conventional transmission line) propagate at all frequencies. Note, however, that TEM modes are only possible when the waveguide possesses an internal conductor running along its length.

10.5 Cylindrical Waves

Consider a cylindrically symmetric wavefunction $\psi(\rho, t)$, where $\rho = \sqrt{x^2 + y^2}$ is a conventional cylindrical polar coordinate. Assuming that this function satisfies the three-dimensional wave equation (10.9), which can be rewritten (see Exercise 3)

$$\frac{\partial^2\psi}{\partial t^2} = v^2\left(\frac{\partial^2\psi}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial\psi}{\partial\rho}\right), \tag{10.29}$$

it is easily demonstrated that a sinusoidal cylindrical wave of phase angle ϕ, wavenumber k, and angular frequency $\omega = k\,v$, takes the form (see Exercise 3)

$$\psi(\rho, t) \simeq \frac{\psi_0}{\rho^{1/2}}\cos(\phi + k\,\rho - \omega\,t) \tag{10.30}$$

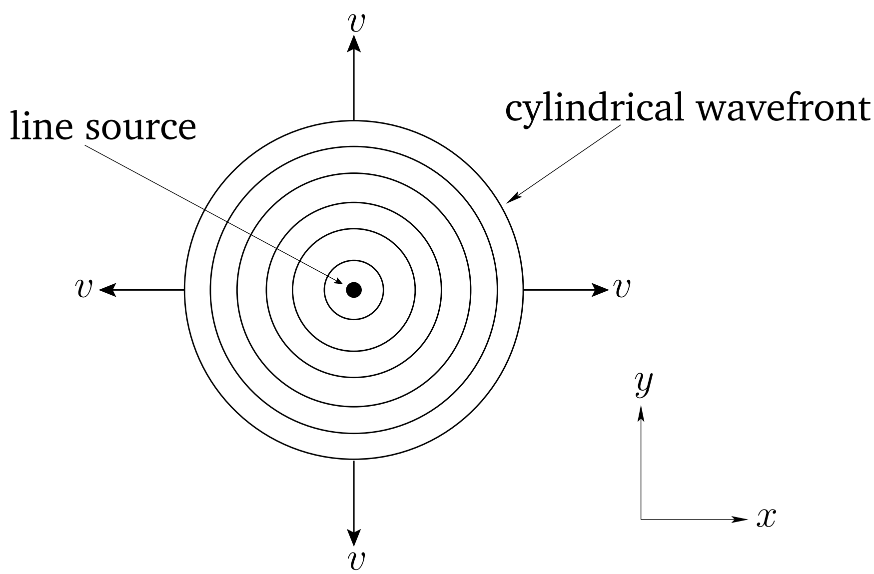


Figure 10.4: *A cylindrical wave.*

in the limit $k\rho \gg 1$. Here, $\psi_0/\rho^{1/2}$ is the amplitude of the wave. In this case, the associated wavefronts (*i.e.*, surfaces of constant phase) are a set of concentric cylinders which propagate radially outward, from their common axis ($\rho = 0$), at the phase velocity $v = \omega/k$. See Figure 10.4. Note that the wave amplitude attenuates as $\rho^{-1/2}$. Such behavior can be understood as a consequence of *energy conservation*, which requires the power flowing across the various $\rho =$ const. surfaces to be constant. (The areas of such surfaces scale as $A \propto \rho$. Moreover, the power flowing across them is proportional to $\psi^2 A$, since the energy flux associated with a wave is generally proportional to ψ^2, and is directed normal to the wavefronts.) The cylindrical wave specified in expression (10.30) is such as would be generated by a uniform *line source* located at $\rho = 0$. See Figure 10.4.

10.6 Exercises

1. Show that the one-dimensional plane wave (10.1) is a solution of the one-dimensional wave equation (10.8) provided that

$$\omega = k\,v.$$

Likewise, demonstrate that the three-dimensional plane wave (10.5) is a solution of the three-dimensional wave equation (10.9) as long as

$$\omega = |\mathbf{k}|\,v.$$

2. Consider a square waveguide of internal dimensions $5 \times 10\,\mathrm{cm}$. What is the frequency (in MHz) of the lowest frequency TE mode which will propagate along the waveguide without attenuation? What are the phase and group velocities (expressed as multiples of c) for a TE mode whose frequency is 5/4 times this cut-off frequency?

3. Demonstrate that for a cylindrically symmetric wavefunction $\psi(\rho, t)$, where $\rho = \sqrt{x^2 + y^2}$, the three-dimensional wave equation (10.9) can be re-written

$$\frac{\partial^2\psi}{\partial t^2} = v^2\left(\frac{\partial^2\psi}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial\psi}{\partial\rho}\right).$$

Show that

$$\psi(\rho, t) \simeq \frac{\psi_0}{\rho^{1/2}}\cos(\phi + k\,\rho - \omega\,t)$$

is an approximate solution of this equation in the limit $k\,\rho \gg 1$, where $v = \omega/k$.

4. Demonstrate that for a spherically symmetric wavefunction $\psi(r, t)$, where $r = \sqrt{x^2 + y^2 + z^2}$, the three-dimensional wave equation (10.9) can be re-written

$$\frac{\partial^2\psi}{\partial t^2} = v^2\left(\frac{\partial^2\psi}{\partial r^2} + \frac{2}{r}\frac{\partial\psi}{\partial r}\right).$$

Show that

$$\psi(r, t) = \frac{\psi_0}{r}\cos(\phi + k\,r - \omega\,t)$$

is a solution of this equation, where $v = \omega/k$. Explain why the wave amplitude attenuates as r^{-1}. What sort of wave source would be most likely to generate the above type of wave solution?

11 Wave Optics

11.1 Introduction

Visible light is a type of *electromagnetic radiation* whose wavelength lies in a relatively narrow band extending from about 400 to 700 nm. The branch of physics which is concerned with the properties of light is known as *optics*. This chapter is devoted to those optical phenomena which depend explicitly on the ultimate *wave* nature of light, and cannot be accounted for using the well-known laws of geometric optics (see Section 10.3). The branch of optics which deals with such phenomena is called *wave optics*. The two most important topics in wave optics are *interference* and *diffraction*. Interference occurs when beams of light from multiple sources (but with similar frequencies), or multiple beams from the same source, intersect one another. Diffraction takes place, for instance, when a single beam of light passes through an opening in an opaque screen whose spatial extent is comparable to the wavelength of the light. It should be noted that interference and diffraction depend on the same underlying physical mechanisms, so that the distinction which is conventionally made between them is somewhat arbitrary.

In the following, for the sake of simplicity, we shall only deal with light emitted from uniform *line sources* interacting with uniform slits which run parallel to these sources, since, under such circumstances, the problem remains essentially *two-dimensional*.

11.2 Two-Slit Interference

Consider a monochromatic plane light wave, propagating in the x-direction, through a transparent dielectric medium of refractive index unity (*e.g.*, a vacuum). (Such a wave might be produced by a uniform line source, running parallel to the z-axis, which is located at $x = -\infty$.) Let the associated wavefunction take the form

$$\psi(x, t) = \psi_0 \cos(\phi + k\,x - \omega\,t). \tag{11.1}$$

Here, ψ represents the *electric* component of the wave, $\psi_0 > 0$ the wave amplitude, ϕ the phase angle, $k > 0$ the wavenumber, $\omega = k\,c$ the angular frequency, and c the velocity of light in vacuum. Let the wave be normally

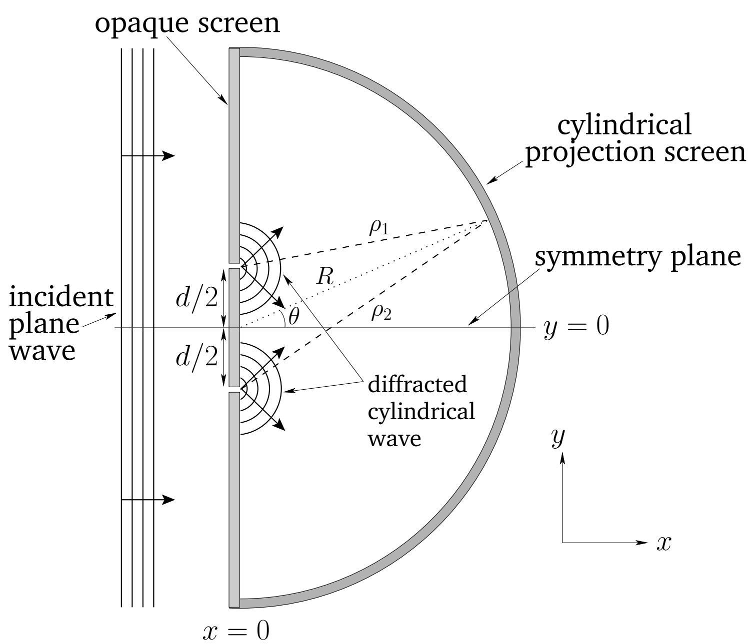


Figure 11.1: *Two-slit interference at normal incidence.*

incident on an opaque screen that is coincident with the plane $x = 0$. See Figure 11.1. Suppose that the screen has two identical slits of width δ cut in it. Let the slits run parallel to the z-axis, be a perpendicular distance d apart, and be located at $y = d/2$ and $y = -d/2$. Suppose that the light which passes through the two slits travels to a cylindrical projection screen of radius R whose axis is the line $x = y = 0$. In the following, it is assumed that there is no variation of wave quantities in the z-direction.

Now, provided that the two slits are much narrower than the wavelength, $\lambda = 2\pi/k$, of the light (*i.e.*, $\delta \ll \lambda$), we expect any light which passes through them to be strongly diffracted. See Section 11.6. *Diffraction* is a fundamental wave phenomenon by which waves bend around small (compared to the wavelength) obstacles, and spread out from narrow (compared to the wavelength) openings, whilst maintaining the same wavelength and frequency. The laws of geometric optics do not take diffraction into account, and are, therefore, restricted to situations in which light interacts with objects whose physical dimensions greatly exceed its wavelength. The assumption of strong diffraction suggests that each slit acts like a uniform *line source* which emits light isotropically in the forward direction (*i.e.*, toward the region $x > 0$), but does not emit light in the backward direction (*i.e.*, toward the region $x < 0$). (It is actually possible to demonstrate that this is, in fact, the case using advanced electromagnetic theory, but such a demonstration lies well beyond the scope of this course.) As discussed in Section 10.5, we would expect a uniform line source to emit a *cylindrical wave*. It follows that each slit emits a half-cylindrical light wave in the forward direction. See Figure 11.1. Moreover, these waves are emitted with *equal amplitude and phase*, since the incident plane wave has the same amplitude (*i.e.*, ψ_0) and phase (*i.e.*, $\phi - \omega\, t$) at both of the slits, and the slits are identical. Finally, we expect the cylindrical waves emitted by the two slits to *interfere* with one another (see Section 7.3), and to, thus, generate a characteristic *interference pattern* on the cylindrical projection screen. Let us now determine the nature of this pattern.

Consider the wave amplitude at a point on the projection screen which lies an angular distance θ from the plane $y = 0$. See Figure 11.1. The wavefunction at this particular point is written

$$\begin{aligned}\psi(\theta, t) \quad \propto \quad & \frac{\cos(\phi + k\,\rho_1 - \omega\, t)}{\rho_1^{1/2}} + \mathcal{O}\left(\frac{1}{k\,\rho_1^{3/2}}\right) \\ & + \frac{\cos(\phi + k\,\rho_2 - \omega\, t)}{\rho_2^{1/2}} + \mathcal{O}\left(\frac{1}{k\,\rho_2^{3/2}}\right), \end{aligned} \tag{11.2}$$

assuming that $k\,\rho_1,\ k\,\rho_2 \gg 1$. In other words, the overall wavefunction in the region $x > 0$ is the superposition of cylindrical wavefunctions [see Equation (10.30)] of equal amplitude (*i.e.*, $\rho^{-1/2}$) and phase (*i.e.*, $\phi + k\,\rho - \omega\, t$) emanating from each slit. Here, ρ_1 and ρ_2 are the distances which the cylindrical waves emitted by the first and second slits (located at $y = d/2$ and $y = -d/2$, respectively) have travelled by the time they reach the point on the projection screen under discussion.

Standard trigonometry (*i.e.*, the law of cosines) reveals that

$$\rho_1 = R\left(1 - \frac{d}{R}\,\sin\theta + \frac{1}{4}\,\frac{d^2}{R^2}\right)^{1/2} = R\left[1 - \frac{1}{2}\,\frac{d}{R}\,\sin\theta + \mathcal{O}\left(\frac{d^2}{R^2}\right)\right]. \tag{11.3}$$

Likewise,

$$\rho_2 = R\left[1 + \frac{1}{2}\,\frac{d}{R}\,\sin\theta + \mathcal{O}\left(\frac{d^2}{R^2}\right)\right]. \tag{11.4}$$

Hence, expression (11.2) yields

$$\begin{aligned}\psi(\theta, t) \quad \propto \quad & \cos(\phi + k\,\rho_1 - \omega\, t) + \cos(\phi + k\,\rho_2 - \omega\, t) \\ & + \mathcal{O}\left(\frac{1}{k\,R}\right) + \mathcal{O}\left(\frac{d}{R}\right), \end{aligned} \tag{11.5}$$

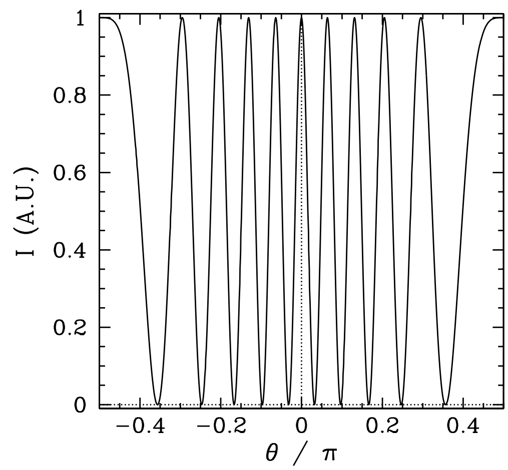


Figure 11.2: *Two-slit far-field interference pattern calculated for* $d/\lambda = 5$ *with normal incidence and narrow slits.*

which, making use of the trigonometric identity $\cos x + \cos y \equiv 2\,\cos[(x + y)/2]\,\cos[(x - y)/2]$, gives

$$\psi(\theta, t) \;\propto\; \cos\left[\phi + \frac{1}{2}\,k\,(\rho_1 + \rho_2) - \omega\,t\right]\cos\left[\frac{1}{2}\,k\,(\rho_1 - \rho_2)\right] + \mathcal{O}\!\left(\frac{1}{k\,R}\right) + \mathcal{O}\!\left(\frac{d}{R}\right), \tag{11.6}$$

or

$$\psi(\theta, t) \;\propto\; \cos\left[\phi + k\,R - \omega\,t + \mathcal{O}\left(\frac{k\,d^2}{R}\right)\right]\cos\left[-\frac{1}{2}\,k\,d\,\sin\theta + \mathcal{O}\left(\frac{k\,d^2}{R}\right)\right] + \mathcal{O}\!\left(\frac{1}{k\,R}\right) + \mathcal{O}\!\left(\frac{d}{R}\right). \tag{11.7}$$

Finally, assuming that

$$\frac{k\,d^2}{R},\ \frac{1}{k\,R},\ \frac{d}{R}, \ll 1, \tag{11.8}$$

the above expression reduces to

$$\psi(\theta, t) \propto \cos(\phi + k\,R - \omega\,t)\,\cos\left(\frac{1}{2}\,k\,d\,\sin\theta\right). \tag{11.9}$$

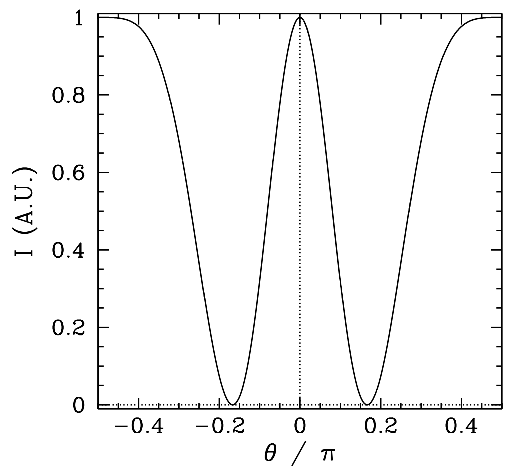


Figure 11.3: *Two-slit far-field interference pattern calculated for* $d/\lambda = 1$ *with normal incidence and narrow slits.*

Now, the orderings (11.8), which can also be written in the form,

$$R \gg d,\, \lambda,\, \frac{d^2}{\lambda}, \tag{11.10}$$

are satisfied provided that the projection screen is located sufficiently far away from the slits. Consequently, the type of interference described in this section is known as *far-field interference*. The characteristic features of far-field interference are that the *amplitudes* of the cylindrical waves emitted by the two slits are approximately equal to one another when they reach a given point on the projection screen (*i.e.*, $|\rho_1 - \rho_2|/\rho_1 \ll 1$), whereas the phases are, in general, significantly different (*i.e.*, $k\,|\rho_1 - \rho_2| \gtrsim \pi$). In other words, the interference pattern generated on the projection screen is entirely due to the *phase difference* between the cylindrical waves emitted by the two slits when they reach the screen. This phase difference is produced by the slight difference in path length between the slits and a given point on the projection screen. (Recall, that the two waves are in phase when they are emitted by the slits.)

The mean energy flux, or *intensity*, of the light striking the projection screen at angular position θ is

$$\begin{aligned}\mathcal{I}(\theta) &\propto \langle \psi(\theta, t)^2 \rangle \\ &\propto \langle \cos^2(\phi + k\,R - \omega\,t)\rangle\, \cos^2\left(\frac{1}{2}\,k\,d\,\sin\theta\right)\end{aligned}$$

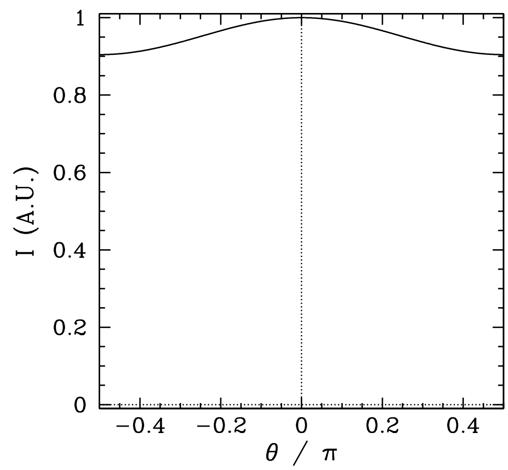


Figure 11.4: *Two-slit far-field interference pattern calculated for* $d/\lambda = 0.1$ *with normal incidence and narrow slits.*

$$\propto \quad \cos^2\left(\frac{1}{2}\,k\,d\,\sin\theta\right), \tag{11.11}$$

where $\langle\cdots\rangle$ denotes an average over a wave period. (The above expression follows from the standard result $\mathcal{I} = E^2/Z_0$, for an electromagnetic wave, where E is the electric component of the wave, and Z_0 the impedance of free space. See Section 7.7. Recall, also, that $\psi \propto E$.) Here, we have made use of the easily established result $\langle\cos^2(\phi + k\,R - \omega\,t)\rangle = 1/2$. Note that, given the very high oscillation frequency of a light wave (*i.e.*, $f \sim 10^{14}$ Hz), it is the intensity of light, rather than the rapidly oscillating amplitude of its electric component, which is typically detected experimentally (*e.g.*, by a photographic film, or photo-multiplier tube). Hence, for the case of two-slit far-field interference, assuming normal incidence and narrow slits, the intensity of the characteristic interference pattern appearing on the projection screen is specified by

$$\mathcal{I}(\theta) \propto \cos^2\left(\pi\,\frac{d}{\lambda}\,\sin\theta\right). \tag{11.12}$$

Figure 11.2 shows the intensity of the typical two-slit far-field interference pattern produced when the slit spacing, d, greatly exceeds the wavelength, λ, of the light. It can be seen that the pattern consists of multiple bright and dark fringes. A bright fringe is generated whenever the cylindrical waves emitted by the two slits interfere constructively at given point on the projection screen. This occurs if the path lengths between the two slits and the point in question differ by an *integer* number of wavelengths: *i.e.*,

$$\rho_2 - \rho_1 = d\,\sin\theta = j\,\lambda, \tag{11.13}$$

where j is an integer, since this ensures that the phases of the two waves differ by an integer multiple of 2π, and, hence, that the effective phase difference is zero. Likewise, a dark fringe is generated whenever the cylindrical waves emitted by the two slits interfere destructively at a given point on the projection screen. This occurs if the path lengths between the two slits and the point in question differ by a *half-integer* number of wavelengths: *i.e.*,

$$\rho_2 - \rho_1 = d\,\sin\theta = (j + 1/2)\,\lambda, \tag{11.14}$$

since this ensures that the effective phase difference between the two waves is π. We conclude that the innermost (*i.e.*, low j, small θ) bright fringes are approximately *equally-spaced*, with a characteristic angular width $\Delta\theta \simeq \lambda/d$. This result, which follows from Equation (11.13), and the small angle approximation $\sin\theta \simeq \theta$, can be used experimentally to determine the wavelength of a monochromatic light source (see Exercise 11.4).

Figure 11.3 shows the intensity of the interference pattern generated when the slit spacing is equal to the wavelength of the light. It can be seen that the width of the central (*i.e.*, $j = 0$, $\theta = 0$) bright fringe has expanded to such an extent that the fringe occupies almost half of the projection screen, leaving room for just two dark fringes on either side of it.

Finally, Figure 11.4 shows the intensity of the interference pattern generated when the slit spacing is much less than the wavelength of the light. It can be seen that the width of the central bright fringe has expanded to such an extent that the band occupies the whole projection screen, and there are no dark fringes. Indeed, $\mathcal{I}(\theta)$ becomes constant in the limit that $d/\lambda \ll 1$, in which case the interference pattern entirely disappears.

It is clear, from Figures 11.2–11.3, that the two-slit far-field interference apparatus shown in Figure 11.1 only generates an interesting interference patten when the slit spacing, d, is greater than the wavelength, λ, of the light.

Suppose, now, that the plane wave which illuminates the interference apparatus is not normally incident on the slits, but instead propagates at an angle θ_0 to the x-axis, as shown in Figure 11.5. In this case, the incident wavefunction (11.1) becomes

$$\psi(x, y, t) = \psi_0\,\cos(\phi + k\,x\,\cos\theta_0 + k\,y\,\sin\theta_0 - \omega\,t). \tag{11.15}$$

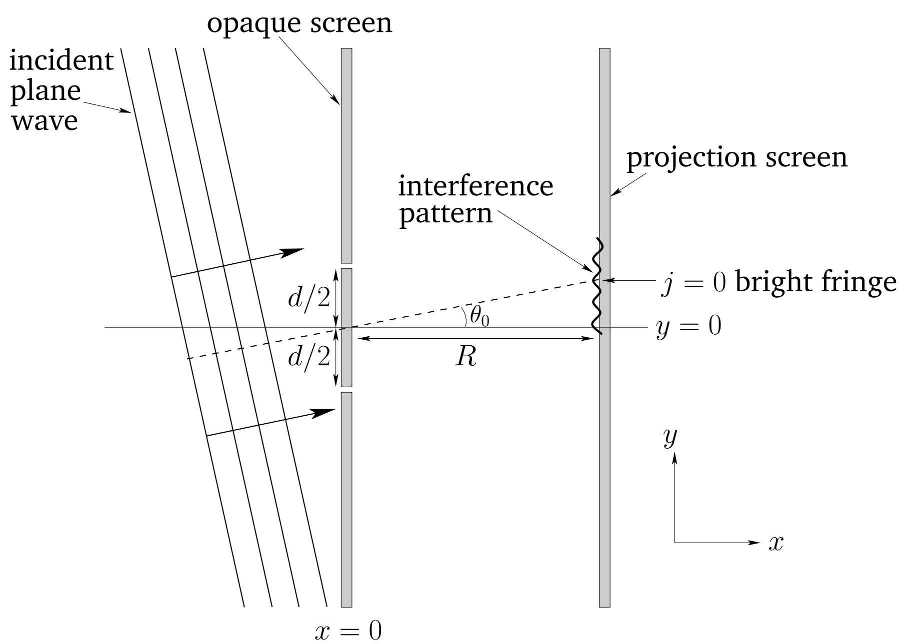


Figure 11.5: *Two-slit interference at oblique incidence.*

Thus, the phase of the light incident on the first slit (located at $x = 0$, $y = d/2$) is $\phi + (1/2)\,k\,d\,\sin\theta_0 - \omega\,t$, whereas the phase of the light incident on the second slit (located at $x = 0$, $y = -d/2$) is $\phi - (1/2)\,k\,d\,\sin\theta_0 - \omega\,t$. Assuming that the cylindrical waves emitted by each slit have the same phase (at the slits) as the plane wave which illuminates them, Equation (11.2) generalizes to

$$\psi(\theta, t) \quad \propto \quad \frac{\cos(\phi_1 + k\,\rho_1 - \omega\,t)}{\rho_1^{1/2}} + \frac{\cos(\phi_2 + k\,\rho_2 - \omega\,t)}{\rho_2^{1/2}}, \tag{11.16}$$

where $\phi_1 = \phi + (1/2)\,k\,d\,\sin\theta_0$ and $\phi_2 = \phi - (1/2)\,k\,d\,\sin\theta_0$. Hence, making use of the far-field orderings (11.10), and a standard trigonometric identity, we obtain

$$\begin{aligned}\psi(\theta, t) \quad &\propto \quad \cos\left[\frac{1}{2}\,(\phi_1 + \phi_2) + k\,R - \omega\,t\right] \cos\left[\frac{1}{2}\,(\phi_1 - \phi_2) - \frac{1}{2}\,k\,d\,\sin\theta\right] \\ &\propto \quad \cos(\phi + k\,R - \omega\,t)\,\cos\left[\frac{1}{2}\,k\,d\,(\sin\theta - \sin\theta_0)\right]. \end{aligned} \tag{11.17}$$

For the sake of simplicity, let us concentrate on the limit $d \gg \lambda$ in which the innermost (*i.e.*, low j) interference fringes are located at small θ. (Note that the projection screen is approximately planar in this limit, as indicated in Figure 11.5, since a sufficiently small section of a cylindrical surface looks like a plane.) Assuming that θ_0 is also small, the above expression reduces to

$$\psi(\theta, t) \propto \cos(\phi + k\,R - \omega\,t)\,\cos\left[\frac{1}{2}\,k\,d\,(\theta - \theta_0)\right], \tag{11.18}$$

and Equation (11.12) becomes

$$\mathcal{I}(\theta) \propto \cos^2\left[\pi\,\frac{d}{\lambda}\,(\theta - \theta_0)\right]. \tag{11.19}$$

Thus, the bright fringes in the interference pattern are located at

$$\theta = \theta_0 + j\,\frac{\lambda}{d}, \tag{11.20}$$

where j is an integer. We conclude that if the slits in a two-slit interference apparatus, such as that shown in Figure 11.5, are illuminated by an obliquely incident plane wave then the consequent phase difference between the cylindrical waves emitted by each slit produces an *angular shift* in the interference pattern appearing on the projection screen. To be more exact, the angular shift is equal to the angle of incidence, θ_0, of the plane wave, so that the central ($j = 0$) bright fringe in the interference pattern is located at $\theta = \theta_0$—see Figure 11.5. Of course, this is equivalent to saying that the position of the central bright fringe can be determined via the rules of geometric optics. (Futhermore, this conclusion holds even when θ_0 is not small.)

11.3 Coherence

A practical monochromatic light source consists of a collection of similar atoms which are continually excited by collisions, and then spontaneously decay back to their electronic ground states, in the process emitting photons of characteristic angular frequency $\omega = \Delta\mathcal{E}/\hbar$, where $\Delta\mathcal{E}$ is the difference in energy between the excited state and the ground state, and $\hbar = 1.055 \times 10^{-34}\,\mathrm{J\,s^{-1}}$ is Planck's constant divided by 2π. Now, an excited electronic state of an atom has a characteristic lifetime, τ, which can be calculated from quantum mechanics, and is typically $10^{-8}\,\mathrm{s}$. It follows that when an atom in an excited state decays back to its ground state it emits a burst of electromagnetic radiation of duration τ and angular frequency ω. However, according to the bandwidth theorem (see Section 8.3), a sinusoidal wave of finite duration τ has a finite bandwidth

$$\Delta\omega \sim \frac{2\pi}{\tau}. \tag{11.21}$$

In other words, if the emitted wave is Fourier transformed in time then it is found to consist of a linear superposition of sinusoidal waves of infinite duration whose frequencies lie in the approximate range $\omega - \Delta\omega/2$ to $\omega + \Delta\omega/2$. We conclude that there is no such thing as a truly monochromatic light source. In reality, all such sources have a small, but finite, bandwidths which are inversely proportional to the lifetimes, τ, of the associated excited atomic states.

So, how do we take the finite bandwidth of a practical "monochromatic" light source into account in our analysis? Actually, all we need to do is to assume that the phase angle, ϕ, appearing in Equations (11.1) and (11.15), is only constant on timescales much less that the lifetime, τ, of the associated excited atomic state, and is subject to abrupt random changes on timescales much greater than τ. We can understand this phenomenon as due to the fact that the radiation emitted by a single atom has a fixed phase angle, ϕ, but only lasts a finite time period, τ, combined with the fact that there is generally no correlation between the phase angles of the radiation emitted by different atoms. Alternatively, we can account for the variation in the phase angle in terms of the finite bandwidth of the light source: *i.e.*, since the light emitted by the source consists of a superposition of sinusoidal waves of frequencies extending over the range $\omega - \Delta\omega/2$ to $\omega + \Delta\omega/2$ then, even if all the component waves start off in phase, the phases will be completely scrambled after a time period $2\pi/\Delta\omega = \tau$ has elapsed. What we are, in effect, saying is that a practical monochromatic light source is *temporally coherent* on timescales much less than its characteristic *coherence time*, τ (which, for visible light, is typically of order 10^{-8} seconds), and *temporally incoherent* on timescales much greater than τ. Incidentally, two waves are said to be *coherent* if their phase difference is constant in time, and *incoherent* if their phase difference varies significantly in time. In this case, the two waves in question are the same wave observed at two different times.

So, what effect does the temporal incoherence of a practical monochromatic light source on timescales greater than $\tau \sim 10^{-8}$ seconds have on the two-slit interference patterns discussed in the previous section? Consider the case of oblique incidence. According to Equation (11.16), the phase angles, $\phi_1 = \phi + (1/2)\, k\, d\, \sin\theta_0$, and $\phi_2 = \phi - (1/2)\, k\, d\, \sin\theta_0$, of the cylindrical waves emitted by each slit are subject to abrupt random changes on

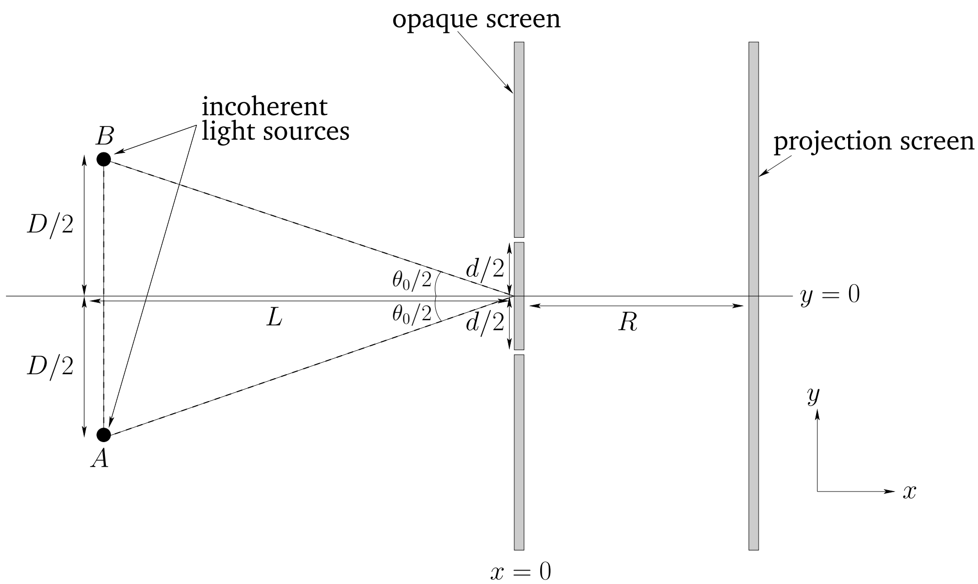


Figure 11.6: *Two-slit interference with two line sources.*

timescales much greater than τ, since the phase angle, ϕ of the plane wave which illuminates the two slits is subject to the same changes. Nevertheless, the *relative phase angle*, $\phi_1 - \phi_2 = k\,d\,\sin\theta_0$, between the two cylindrical waves remains *constant*. Moreover, as is clear from Equation (11.17), the interference pattern appearing on the projection screen is generated by the *phase difference* $(1/2)\,(\phi_1 - \phi_2) - (1/2)\,k\,d\,\sin\theta$ between the two cylindrical waves at a given point on the screen, and this phase difference only depends on the relative phase angle. Indeed, the intensity of the interference pattern is $\mathcal{I}(\theta) \propto \cos^2[(1/2)\,(\phi_1 - \phi_2) - (1/2)\,k\,d\,\sin\theta]$. Hence, the fact that the relative phase angle, $\phi_1 - \phi_2$, between the two cylindrical waves emitted by the slits remains constant on timescales much longer than the characteristic coherence time, τ, of the light source implies that the interference pattern generated in a conventional two-slit interference apparatus is generally *unaffected* by the temporal incoherence of the source. Strictly speaking, however, the preceding conclusion is only accurate when the spatial extent of the light source is *negligible*. Let us now broaden our discussion to take spatially extended light sources into account.

Up until now, we have assumed that our two-slit interference apparatus is illuminated by a single plane wave, such as might be generated by a line source located at infinity. Let us now consider a more realistic situation in which the light source is located a finite distance from the slits, and also has a finite spatial extent. Figure 11.6 shows the simplest possible case. Here, the slits are illuminated by two identical line sources, A and B, which are a distance D apart, and a perpendicular distance L from the opaque screen containing the slits. Assuming that $L \gg D$, d, the light incident on the slits from source A is effectively a plane wave whose direction of propagation subtends an angle $\theta_0/2 \simeq D/2\,L$ with the x-axis. Likewise, the light incident on the slits from source B is a plane wave whose direction of propagation subtends an angle $-\theta_0/2$ with the x-axis. Moreover, the net interference pattern (*i.e.*, wavefunction) appearing on the projection screen is the *linear superposition* of the patterns generated by each source taken individually (since light propagation is ultimately governed by a *linear* wave equation with *superposable* solutions—see Section 10.2.). Let us determine whether these patterns reinforce one another, or interfere with one another.

The light emitted by source A has a phase angle, ϕ_A, which is constant on timescales much less than the characteristic coherence time of the source, τ, but is subject to abrupt random changes on timescale much longer than τ. Likewise, the light emitted by source B has a phase angle, ϕ_B, which is constant on timescales much less than τ, and varies significantly on timescales much greater than τ. Furthermore, there is, in general, *no correlation* between ϕ_A and ϕ_B. In other words, our composite light source, consisting of the two line sources A and B, is both *temporally* and *spatially* incoherent on timescales much longer than τ.

Again working in the limit $d \gg \lambda$, with θ, $\theta_0 \ll 1$, Equation (11.18) yields the following expression for the wavefunction on the projection screen:

$$\begin{aligned}\psi(\theta,t) &\propto \cos(\phi_A + k\,R - \omega\,t)\,\cos\left[\frac{1}{2}\,k\,d\,(\theta-\theta_0/2)\right]\\ &\quad+\cos(\phi_B + k\,R - \omega\,t)\,\cos\left[\frac{1}{2}\,k\,d\,(\theta+\theta_0/2)\right].\end{aligned} \tag{11.22}$$

Hence, the intensity of the interference pattern is

$$\begin{aligned}\mathcal{I}(\theta) \propto \langle\psi^2\rangle &\propto \langle\cos^2(\phi_A + k\,R - \omega\,t)\rangle\cos^2\left[\frac{1}{2}\,k\,d\,(\theta-\theta_0/2)\right]\\ &\quad+2\,\langle\cos(\phi_A + k\,R - \omega\,t)\,\cos(\phi_B + k\,R - \omega\,t)\rangle\\ &\quad\times\cos\left[\frac{1}{2}\,k\,d\,(\theta-\theta_0/2)\right]\,\cos\left[\frac{1}{2}\,k\,d\,(\theta+\theta_0/2)\right]\\ &\quad+\langle\cos^2(\phi_B + k\,R - \omega\,t)\rangle\cos^2\left[\frac{1}{2}\,k\,d\,(\theta+\theta_0/2)\right].\end{aligned}$$

(11.23)

However, $\langle\cos^2(\phi_A + k\,R - \omega\,t)\rangle = \langle\cos^2(\phi_B + k\,R - \omega\,t)\rangle = 1/2$, and $\langle\cos(\phi_A + k\,R - \omega\,t)\,\cos(\phi_B + k\,R - \omega\,t)\rangle = 0$, since the phase angles ϕ_A and ϕ_B are uncorrelated. Hence, the above expression reduces to

$$\begin{aligned}\mathcal{I}(\theta) &\propto \cos^2\left[\frac{1}{2}\,k\,d\,(\theta - \theta_0/2)\right] + \cos^2\left[\frac{1}{2}\,k\,d\,(\theta + \theta_0/2)\right]\\ &= 1 + \cos\left(2\pi\,\frac{d}{\lambda}\,\theta\right)\,\cos\left(\pi\,\frac{d}{\lambda}\,\theta_0\right), \qquad (11.24)\end{aligned}$$

where use has been made of the trigonometric identities $\cos^2\theta \equiv (1 + \cos 2\theta)/2$, and $\cos x + \cos y \equiv 2\,\cos[(x + y)/2]\,\cos((x - y)/2]$. Note that if $\theta_0 = \lambda/2\,d$ then $\cos[\pi\,(d/\lambda)\,\theta_0] = 0$ and $\mathcal{I}(\theta) \propto 1$. In this case, the bright fringes of the interference pattern generated by source A exactly overlay the dark fringes of the pattern generated by source B, and *vice versa*, and the net interference pattern is completely washed out. On the other hand, if $\theta_0 \ll \lambda/d$ then $\cos[\pi\,(d/\lambda)\,\theta_0] = 1$ and $\mathcal{I}(\theta) \propto 1 + \cos[2\pi\,(d/\lambda)\,\theta] = 2\,\cos^2[\pi\,(d/\lambda)\,\theta]$. In this case, the two interference patterns reinforce one another, and the net interference pattern is the same as that generated by a light source of negligible spatial extent.

Suppose, now, that our light source consists of a regularly spaced array of very many identical incoherent line sources, filling the region between sources A and B in Figure 11.6. In other words, suppose that our light source is a uniform incoherent source of angular extent θ_0. It is, hopefully, clear, from the linear nature of the problem, that the associated interference pattern can be obtained by *averaging* expression (11.24) over all θ_0 values in the range 0 to θ_0: *i.e.*, by operating on this expression with $\theta_0^{-1}\int_0^{\theta_0}\cdots\,d\theta_0$. In this manner, we obtain

$$\mathcal{I}(\theta) \propto 1 + \cos\left(2\pi\,\frac{d}{\lambda}\,\theta\right)\,\mathrm{sinc}\left(\pi\,\frac{d}{\lambda}\,\theta_0\right), \qquad (11.25)$$

where $\mathrm{sinc}(x) \equiv \sin x/x$. Now, we can conveniently parameterize the *visibility* of the interference pattern, appearing on the projection screen, in terms of the quantity

$$V = \frac{\mathcal{I}_{\max} - \mathcal{I}_{\min}}{\mathcal{I}_{\max} + \mathcal{I}_{\min}}, \qquad (11.26)$$

where the maximum and minimum values of the intensity are taken with respect to variation in θ (rather than θ_0). Of course, $V = 1$ corresponds to a

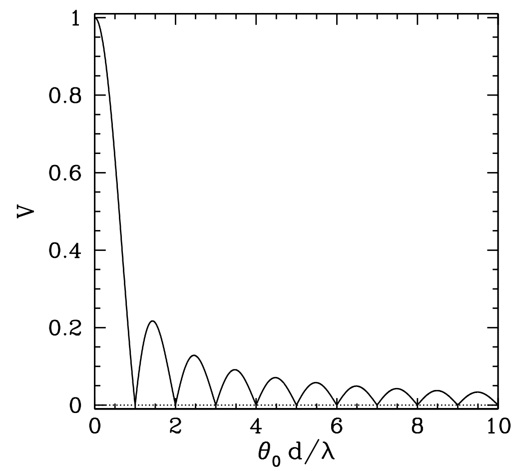


Figure 11.7: *Visibility of a two-slit far-field interference pattern generated by an extended incoherent light source.*

sharply defined pattern, and $V = 0$ to a pattern which is completely washed out. It follows from (11.25) that

$$V = \left|\mathrm{sinc}\left(\pi\,\frac{d}{\lambda}\,\theta_0\right)\right|. \tag{11.27}$$

The predicted visibility, V, of a two-slit interference pattern generated by an extended incoherent light source is plotted as a function of the angular extent, θ_0, of the source in Figure 11.7. It can be seen that the pattern is highly visible (*i.e.*, $V \sim 1$) when $\theta_0 \ll \lambda/d$, but becomes washed out (*i.e.*, $V \sim 0$) when $\theta_0 \gtrsim \lambda/d$.

We conclude that a spatially extended incoherent light source only generates a visible interference pattern in a conventional two-slit interference apparatus when the angular extent of the source is sufficiently small: *i.e.*, when

$$\theta_0 \ll \frac{\lambda}{d}. \tag{11.28}$$

Equivalently, if the source is of linear extent D, and located a distance L from the slits, then the source only generates a visible interference pattern when it is sufficiently far away from the slits: *i.e.*, when

$$L \gg \frac{d\,D}{\lambda}. \tag{11.29}$$

This follows because $\theta_0 \simeq D/L$.

The whole of the above discussion is premised on the assumption that an extended light source is both temporally and spatially incoherent on timescales much longer than a typical atomic coherence time, which is about 10^{-8} seconds. This is, indeed, generally the case. However, there is one type of light source—namely, a *laser*—for which this is not necessarily the case. In a laser (in single-mode operation), excited atoms are stimulated in such a manner that they emit radiation which is both temporally and spatially coherent on timescales much longer than the relevant atomic coherence time.

Let us, briefly, consider the two-slit far-field interference pattern generated by an extended *coherent* light source. Suppose that the two line sources, A and B, in Figure 11.6 are mutually coherent (*i.e.*, $\phi_A = \phi_B$). In this case, as is easily demonstrated, Equation (11.24) is replaced by

$$\begin{aligned} \mathcal{I}(\theta) &\propto \left(\cos\left[\frac{1}{2}\,k\,d\,(\theta-\theta_0/2)\right] + \cos\left[\frac{1}{2}\,k\,d\,(\theta+\theta_0/2)\right]\right)^2 \\ &= 4\,\cos^2\left(\pi\,\frac{d}{\lambda}\,\theta\right)\,\cos^2\left(\frac{\pi}{2}\,\frac{d}{\lambda}\,\theta_0\right) \\ &= 2\,\cos^2\left(\pi\,\frac{d}{\lambda}\,\theta\right)\left[1+\cos\left(\pi\,\frac{d}{\lambda}\,\theta_0\right)\right]. \end{aligned} \tag{11.30}$$

Moreover, when this expression is averaged over θ_0, in order to generate the interference pattern produced by a uniform coherent light source of angular extent θ_0, we obtain

$$\mathcal{I}(\theta) \propto 2\,\cos^2\left(\pi\,\frac{d}{\lambda}\,\theta\right)\left[1+\mathrm{sinc}\left(\pi\,\frac{d}{\lambda}\,\theta_0\right)\right]. \tag{11.31}$$

It follows, from (11.26), that the visibility of the interference pattern is *unity*: *i.e.*, the pattern is sharply defined irrespective of the angular extent, θ_0, of the light source, as long as the source is spatially coherent. It is hardly surprising, then, that lasers generally produce much clearer interference patterns than conventional incoherent light sources.

11.4 Multi-Slit Interference

Suppose that the interference apparatus pictured in Figure 11.1 is modified such that N identical slits of width $\delta \ll \lambda$, running parallel to the z-axis, are cut in the opaque screen which occupies the plane $x = 0$. Let the slits be located at $y = y_n$, for $n = 1, N$. For the sake of simplicity, the arrangement of slits is assumed to be *symmetric* with respect to the plane $y = 0$: *i.e.*, if there is a slit at $y = y_n$ then there is also a slit at $y = -y_n$. Now, the path length between a point on the projection screen which is an angular distance θ from the plane $y = 0$ and the nth slit is [*cf.*, Equation (11.3)]

$$\rho_n = R\left[1 - \frac{y_n}{R}\,\sin\theta + \mathcal{O}\left(\frac{y_n^2}{R^2}\right)\right]. \tag{11.32}$$

Thus, making use of the far-field orderings (11.10), where d now represents the typical spacing between neighboring slits, and assuming normally incident collimated light, Equation (11.5) generalizes to

$$\psi(\theta, t) \propto \sum_{n=1,N} \cos(\phi + k\,R - \omega\,t - k\,y_n\,\sin\theta), \tag{11.33}$$

which can also be written

$$\begin{aligned}\psi(\theta, t) \quad &\propto \quad \cos(\phi + k\,R - \omega\,t) \sum_{n=1,N} \cos(k\,y_n\,\sin\theta) \\ &\qquad + \sin(\phi + k\,R - \omega\,t) \sum_{n=1,N} \sin(k\,y_n\,\sin\theta),\end{aligned} \tag{11.34}$$

or

$$\psi(\theta, t) \propto \cos(\phi + k\,R - \omega\,t) \sum_{n=1,N} \cos(k\,y_n\,\sin\theta). \tag{11.35}$$

Here, we have made use of the fact that arrangement of slits is symmetric with respect to the plane $y = 0$ (which implies that $\sum_{n=1,N} \sin(k\,y_n\,\sin\theta) = 0$). We have also employed the standard identity $\cos(x - y) \equiv \cos x\,\cos y + \sin x\,\sin y$. It follows that the intensity of the interference pattern appearing on the projection screen is specified by

$$\mathcal{I}(\theta) \propto \langle\psi(\theta, t)^2\rangle \propto \left[\sum_{n=1,N} \cos\left(2\pi\,\frac{y_n}{\lambda}\,\sin\theta\right)\right]^2, \tag{11.36}$$

since $\langle\cos^2(\phi + k\,R - \omega\,t)\rangle = 1/2$. The above expression is a generalized version of Equation (11.12).

Suppose that the slits are *evenly spaced* a distance d apart, so that

$$y_n = [n - (N+1)/2]\,d \tag{11.37}$$

for $n = 1, N$. It follows that

$$\mathcal{I}(\theta) \propto \left[\sum_{n=1,N} \cos\left(2\pi\,[n - (N+1)/2]\,\frac{d}{\lambda}\,\sin\theta\right)\right]^2, \tag{11.38}$$

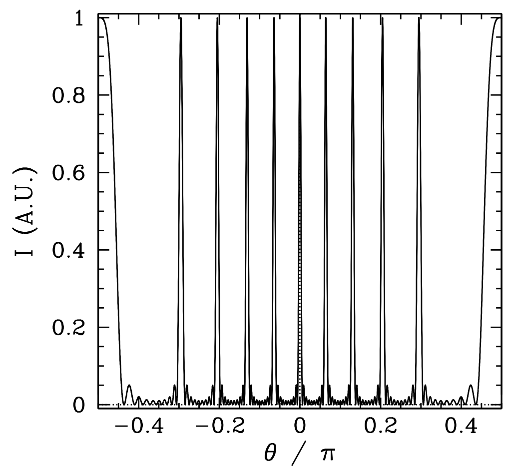


Figure 11.8: *Multi-slit far-field interference pattern calculated for* $N = 10$ *and* $d/\lambda = 5$ *with normal incidence and narrow slits.*

which can be summed to give (see Exercise 11.1)

$$\mathcal{I}(\theta) \propto \frac{\sin^2[\pi\, N\,(d/\lambda)\,\sin\theta]}{\sin^2[\pi\,(d/\lambda)\,\sin\theta)]}. \tag{11.39}$$

Now, the multi-slit interference function, (11.39), exhibits strong maxima in situations in which its numerator and denominator are *simultaneously zero*: *i.e.*, when

$$\sin\theta = j\,\frac{\lambda}{d}, \tag{11.40}$$

where j is an integer. In this situation, application of *L'Hopital's rule* yields $\mathcal{I} = N^2$. The heights of these so-called *principal maxima* in the interference function are very large, being proportional to N^2, because there is constructive interference of the light from *all* N slits. This occurs because the path lengths between neighboring slits and the point on the projection screen at which a given maximum is located differ by an *integer* number of wavelengths: *i.e.*, $\rho_n - \rho_{n-1} = d\,\sin\theta = j\,\lambda$. Note, incidentally, that all of the principle maxima have the *same* height.

The multi-slit interference function (11.39) is *zero* when its numerator is zero but its denominator non-zero: *i.e.*, when

$$\sin\theta = \frac{l}{N}\,\frac{\lambda}{d}, \tag{11.41}$$

where l is an integer which is *not* an integer multiple of N. It follows that there are $N-1$ zeros between neighboring principle maxima. It can also be demonstrated that there are $N-2$ *secondary maxima* between the said zeros. However, these maxima are much lower in height, by a factor of order N^2, than the primary maxima.

Figure 11.8 shows the typical far-field interference pattern produced by a system of ten identical equally-spaced parallel slits, assuming normal incidence and narrow slits, when the slit spacing, d, greatly exceeds the wavelength, λ, of the light (which, as we saw in Section 11.2, is the most interesting case). It can be seen that the pattern consists of a series of bright fringes of equal height, separated by much wider (relatively) dark fringes. Of course, the bright fringes correspond to the principal maxima discussed above. As is the case for two-slit interference, the innermost (*i.e.*, low j, small θ) principal maxima are approximately *equally-spaced*, with a characteristic angular spacing $\Delta\theta \simeq \lambda/d$. [This result follows from Equation (11.40), and the small angle approximation $\sin\theta \simeq \theta$.] However, the typical angular width of a principal maximum (*i.e.*, the angular distance between the maximum and the closest zeroes on either side of it) is $\delta\theta \simeq (1/N)(\lambda/d)$. [This result follows from Equation (11.41), and the small angle approximation]. The ratio of the angular width of a principal maximum to the angular spacing between successive maxima is thus

$$\frac{\delta\theta}{\Delta\theta} \simeq \frac{1}{N}. \tag{11.42}$$

Hence, we conclude that, as the number of slits increases, the bright fringes in a multi-slit interference pattern become progressively sharper.

The most common practical application of multi-slit interference is the *transmission diffraction grating*. Such a device consists of N identical equally-spaced parallel scratches on one side of a thin uniform transparent glass or plastic film. When the film is illuminated the scratches strongly scatter the incident light, and effectively constitute N identical equally-spaced parallel line sources. Hence, the grating generates the type of N-slit interference pattern discussed above, with one major difference: *i.e.*, the central ($j=0$) principal maximum has contributions not only from the scratches, but also from all the transparent material between the scratches. Thus, the central principal maximum is considerably brighter than the other ($j \neq 0$) principal maxima.

Diffraction gratings are often employed in *spectroscopes*, which are instruments used to decompose light that is made up of a mixture of different wavelengths into its constituent wavelengths. As a simple example, suppose that a spectroscope contains an N-line diffraction grating which is illuminated, at normal incidence, by a mixture of light of wavelength λ, and light of wavelength $\lambda + \Delta\lambda$, where $\Delta\lambda \ll \lambda$. As always, the overall interference pattern (*i.e.*, the overall wavefunction at the projection screen) produced by the grating is a *linear superposition* of the pattern generated by the light of wavelength λ, and the pattern generated by the light of wavelength $\lambda + \Delta\lambda$. Consider the jth-order principal maximum associated with the wavelength λ interference pattern, which is located at θ_j, where $\sin\theta_j = j\,(\lambda/d)$—see Equation (11.40). Here, d is the spacing between neighboring lines on the diffraction grating, which is assumed to be greater than λ. (Incidentally, the width of the lines is assumed to be much less than λ.) Of course, the maximum in question has a finite angular width. We can determine this width by locating the zeros in the interference pattern on either side of the maximum. Let the zeros be located at $\theta_j \pm \delta\theta_j$. Now, the maximum itself corresponds to $\pi\,N\,(d/\lambda)\,\sin\theta_j = \pi\,N\,j$. Hence, the zeros correspond to $\pi\,N\,(d/\lambda)\,\sin(\theta_j \pm \delta\theta_j) = \pi\,(N\,j \pm 1)$ [*i.e.*, they correspond to the first zeros of the function $\sin[\pi\,N\,(d/\lambda)\,\sin\theta]$ on either side of the zero at θ_j—see Equation (11.39)]. Taylor expanding to first-order in $\delta\theta_j$, we obtain

$$\delta\theta_j = \frac{\tan\theta_j}{N\,j}. \tag{11.43}$$

Hence, the maximum in question effectively extends from $\theta_j - \delta\theta_j$ to $\theta_j + \delta\theta_j$. Consider, now, the jth-order principal maximum associated with the wavelength $\lambda + \Delta\lambda$ interference pattern, which is located at $\theta_j + \Delta\theta_j$, where $\sin(\theta_j + \Delta\theta_j) = j\,(\lambda + \Delta\lambda)/d$—see Equation (11.40). Taylor expanding to first-order in $\Delta\theta_j$, we obtain

$$\Delta\theta_j = \tan\theta_j\,\frac{\Delta\lambda}{\lambda}. \tag{11.44}$$

Clearly, in order for the spectroscope to resolve the incident light into its two constituent wavelengths, at the jth spectral order, the angular spacing, $\Delta\theta_j$, between the jth-order maxima associated with these two wavelengths must be greater than the angular widths, $\delta\theta_j$, of the maxima themselves. If this is the case then the overall jth-order maximum will consist of two closely spaced maxima, or "spectral lines" (centered at θ_j and $\theta_j + \Delta\theta_j$). On the other hand, if this is not the case then the two maxima will merge to form a single maximum, and it will consequently not be possible to tell that the incident light consists of a mixture of *two* different wavelengths. Thus, the condition for the spectroscope to be able to resolve the spectral lines at the jth spectral order is $\Delta\theta_j > \delta\theta_j$, or

$$\frac{\Delta\lambda}{\lambda} > \frac{1}{N\,j}. \tag{11.45}$$

We conclude that the resolving power of a diffraction grating spectroscope increases as the number of illuminated lines (*i.e.*, N) increases, and also as the spectral order (*i.e.*, j) increases. Note, incidentally, that there is no resolving power at the lowest (*i.e.*, $j = 0$) spectral order, since the corresponding principal maximum is located at $\theta = 0$ irrespective of the wavelength of the incident light. Moreover, there is a limit to how large j can become (*i.e.*, a given diffraction grating, illuminated by light of a given wavelength, has a finite number of principal maxima). This follows since $\sin\theta_j$ cannot exceed unity, so, according to (11.40), j cannot exceed d/λ.

11.5 Fourier Optics

Up to now, we have only considered the interference of monochromatic light produced when a plane wave is incident on an opaque screen, coincident with the plane $x = 0$, which has a number of *narrow* (*i.e.*, $\delta \ll \lambda$, where δ is the slit width) slits, running parallel to the z-axis, cut in it. Let us now generalize our analysis to take slits of *finite width* (*i.e.*, $\delta \gtrsim \lambda$) into account. In order to achieve this goal, it is convenient to define the so-called *aperture function*, $F(y)$, of the screen. This function takes the value zero if the screen is opaque at position y, and some constant positive value if it is transparent, and is normalized such that $\int_{-\infty}^{\infty} F(y)\,dy = 1$. Thus, for the case of a screen with N identical slits of negligible width, located at $y = y_n$, for $n = 1, N$, the appropriate aperture function is

$$F(y) = \frac{1}{N}\sum_{y=1,N}\delta(y - y_n), \tag{11.46}$$

where $\delta(y)$ is a Dirac delta function.

The wavefunction at the projection screen, generated by the above mentioned arrangement of slits, when the opaque screen is illuminated by a plane wave of phase angle ϕ, wavenumber k, and angular frequency ω, whose direction of propagation subtends an angle θ_0 with the x-axis, is (see the analysis in Sections 11.2 and 11.4)

$$\psi(\theta, t) \propto \cos(\phi + k\,R - \omega\,t)\sum_{n=1,N}\cos[k\,y_n\,(\sin\theta - \sin\theta_0)]. \tag{11.47}$$

Here, for the sake of simplicity, we have assumed that the arrangement of slits is symmetric with respect to the plane $y = 0$, so that $\sum_{n=1,N} \sin(\alpha\, y_n) = 0$ for any α. Now, using the well-known properties of the delta function [see Equation (8.26)], expression (11.47) can also be written

$$\psi(\theta, t) \propto \cos(\phi + k\,R - \omega\,t)\,\bar{F}(\theta), \tag{11.48}$$

where

$$\bar{F}(\theta) = \int_{-\infty}^{\infty} F(y)\ \cos[k\,(\sin\theta - \sin\theta_0)\,y]\,dy. \tag{11.49}$$

In the following, we shall assume that Equation (11.48) is a *general result*, and is valid even when the slits in the opaque screen are of finite width (*i.e.*, $\delta \gtrsim \lambda$). This assumption is equivalent to the assumption that each unblocked section of the screen emits a cylindrical wave in the forward direction that is in phase with the plane wave which illuminates it from behind. The latter assumption is known as *Huygen's principle*. (Huygen's principle can actually be justified using advanced electromagnetic theory, but such a proof lies well beyond the scope of this course.) Note that the *interference/diffraction function*, $\bar{F}(\theta)$, is just the *Fourier transform* of the aperture function, $F(y)$. This is an extremely powerful result. It implies that we can work out the far-field interference/diffraction pattern associated with *any* arrangement of parallel slits, of arbitrary width, by simply Fourier transforming the associated aperture function. Of course, once we have calculated the interference/diffraction function, $\bar{F}(\theta)$, the intensity of the interference/diffraction pattern appearing on the projection screen is readily obtained from

$$\mathcal{I}(\theta) \propto \left[\bar{F}(\theta)\right]^2. \tag{11.50}$$

11.6 Single-Slit Diffraction

Suppose that the opaque screen contains a single slit of finite width. In fact, let the slit in question be of width δ, and extend from $y = -\delta/2$ to $y = \delta/2$. Thus, the associated aperture function is

$$F(y) = \left\{ \begin{array}{lll} 1/\delta & & |y| \leq \delta/2 \\ 0 & & |y| > \delta/2 \end{array} \right. . \tag{11.51}$$

It follows from (11.49) that

$$\bar{F}(\theta) = \frac{1}{\delta} \int_{-\delta/2}^{\delta/2} \cos[k\,(\sin\theta - \sin\theta_0)\,y]\,dy = \text{sinc}\left[\pi\,\frac{\delta}{\lambda}\,(\sin\theta - \sin\theta_0)\right], \tag{11.52}$$

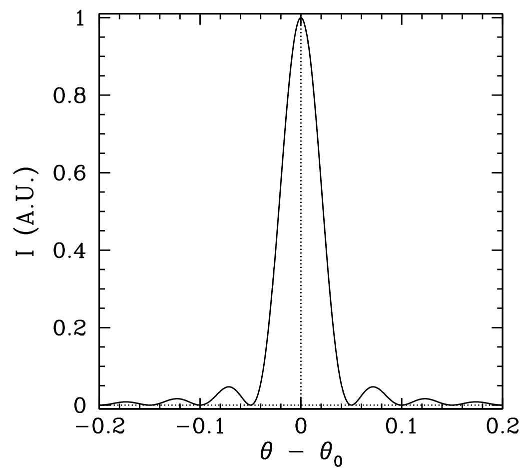


Figure 11.9: *Single-slit far-field diffraction pattern calculated for* $\delta/\lambda = 20$.

where $\mathrm{sinc}(x) \equiv \sin(x)/x$. Finally, assuming, for the sake of simplicity, that θ, $\theta_0 \ll 1$, which is most likely to be the case when $\delta \gg \lambda$, the diffraction pattern appearing on the projection screen is specified by

$$\mathcal{I}(\theta) \propto \mathrm{sinc}^2\left[\pi\,\frac{\delta}{\lambda}\,(\theta-\theta_0)\right]. \tag{11.53}$$

Note, from *L'Hopital's rule*, that $\mathrm{sinc}(0) = \lim_{x\to 0}\,\sin x/x = \lim_{x\to 0}\,\cos x/1 = 1$. Furthermore, it is easily demonstrated that the zeros of the $\mathrm{sinc}(x)$ function occur at $x = j\,\pi$, where j is a *non-zero* integer.

Figure 11.9 shows a typical single-slit diffraction pattern calculated for a case in which the slit width greatly exceeds the wavelength of the light. Note that the pattern consists of a dominant central maximum, flanked by subsidiary maxima of negligible amplitude. The situation is shown schematically in Figure 11.10. When the incident plane wave, whose direction of motion subtends an angle θ_0 with the x-axis, passes through the slit it is effectively transformed into a *divergent* beam of light (where the beam corresponds to the central peak in Figure 11.9) that is centered on $\theta = \theta_0$. The angle of divergence of the beam, which is obtained from the first zero of the single-slit diffraction function (11.53), is

$$\delta\theta = \frac{\lambda}{\delta}: \tag{11.54}$$

i.e., the beam effectively extends from $\theta_0 - \delta\theta$ to $\theta_0 + \delta\theta$. Thus, if the slit width, δ, is very much greater than the wavelength, λ, of the light then the

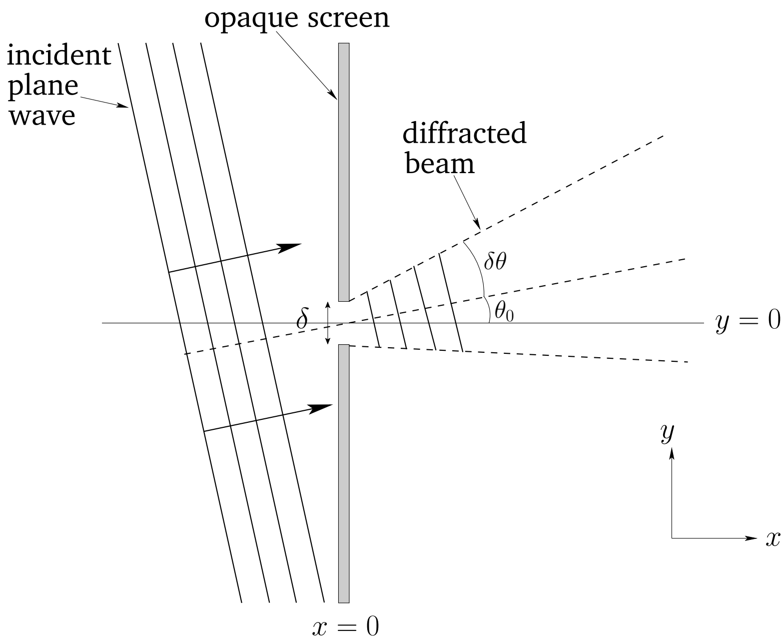


Figure 11.10: *Single-slit diffraction at oblique incidence.*

beam divergence is negligible, and the beam is, thus, governed by the laws of geometric optics (according to which there is no beam divergence). On the other hand, if the slit width is comparable with the wavelength of the light then the beam divergence is significant, and the behavior of the beam is, consequently, very different than that predicted by the laws of geometric optics.

The diffraction of light is a very important physical phenomenon, since it sets a limit on the angular resolution of optical instruments. For instance, consider a telescope whose objective lens is of diameter D. When a plane wave from a distant light source of negligible angular extent (*e.g.*, a star) enter the lens it is diffracted, and forms a divergent beam of angular width λ/D. Thus, instead of being a point, the resulting image of the star is a *disk* of finite angular width λ/D. Such a disk is known as an *Airy disk*. Suppose that two stars are an angular distance $\Delta\theta$ apart in the sky. As we have just seen, when viewed through the telescope, each star appears as a disk of angular extent $\delta\theta = \lambda/D$. Clearly, if $\Delta\theta > \delta\theta$ then the two stars appear as separate disks. On the other hand, if $\Delta\theta < \delta\theta$ then the two disks merge to form a single disk, and it becomes impossible to tell that there are, in fact, *two* stars. It follows that the maximum angular resolution of a telescope whose objective lens is of diameter D is

$$\delta\theta \simeq \frac{\lambda}{D}. \tag{11.55}$$

This result is usually called the *Rayleigh criterion*. Note that the angular resolution of the telescope increases (*i.e.*, $\delta\theta$ decreases) as the diameter of its objective lens increases.

11.7 Multi-Slit Diffraction

Suppose, finally, that the opaque screen in our interference/diffraction apparatus contains N identical equally-spaced parallel slits of *finite width*. Let the slit spacing be d, and the slit width δ, where $\delta < d$. It follows that the aperture function for the screen is written

$$F(y) = \frac{1}{N} \sum_{n=1,N} F_2(y - y_n), \tag{11.56}$$

where

$$y_n = [n - (N+1)/2]\, d, \tag{11.57}$$

and

$$F_2(y) = \left\{ \begin{array}{lll} 1/\delta & & |y| \leq \delta/2 \\ 0 & & |y| > \delta/2 \end{array} \right. . \tag{11.58}$$

Of course, $F_2(y)$ is the aperture function for a single slit, of finite width δ, which is centered on $\theta = 0$—see Equation (11.51).

Assuming normal incidence (*i.e.*, $\theta_0 = 0$), the interference/diffraction function, which is the Fourier transform of the aperture function, takes the form [see Equation (11.49)]

$$\bar{F}(\theta) = \int_{-\infty}^{\infty} F(y)\, \cos(k\, \sin\theta\, y)\, dy. \tag{11.59}$$

Hence,

$$\begin{aligned} \bar{F}(\theta) &= \frac{1}{N} \sum_{n=1,N} \int_{-\infty}^{\infty} F_2(y - y_n)\, \cos(k\, \sin\theta\, y)\, dy \\ &= \frac{1}{N} \sum_{n=1,N} \left[\cos(k\, \sin\theta\, y_n) \int_{-\infty}^{\infty} F_2(y')\, \cos(k\, \sin\theta\, y')\, dy' \right. \end{aligned}$$

$$
\begin{aligned}
&\quad - \sin(k\,\sin\theta\,y_n)\int_{-\infty}^{\infty} F_2(y')\,\sin(k\,\sin\theta\,y')\,dy'\Bigg] \qquad (11.60)\\
&= \left[\frac{1}{N}\sum_{n=1,N}\cos(k\,\sin\theta\,y_n)\right]\int_{-\infty}^{\infty} F_2(y')\,\cos(k\,\sin\theta\,y')\,dy',
\end{aligned}
$$

where $y' = y - y_n$. Here, we have made use of $\int_{-\infty}^{\infty} F_2(y')\,\sin(\alpha\,y')\,dy' = 0$, for any α, which follows because $F_2(y')$ is even in y', whereas $\sin(\alpha\,y')$ is odd. We have also employed the standard identity $\cos(x-y) \equiv \cos x\,\cos y - \sin x\,\sin y$. The above expression reduces to

$$\bar{F}(\theta) = \bar{F}_1(\theta)\,\bar{F}_2(\theta). \qquad (11.61)$$

Here [*cf.*, Equation (11.39)],

$$
\begin{aligned}
\bar{F}_1(\theta) &= \int_{-\infty}^{\infty} F_1(y)\,\cos(k\,\sin\theta\,y)\,dy = \frac{1}{N}\sum_{n=1,N}\cos(k\,\sin\theta\,y_n)\\
&= \frac{1}{N}\,\frac{\sin[\pi\,N\,(d/\lambda)\,\sin\theta]}{\sin[\pi\,(d/\lambda)\,\sin\theta]}, \qquad (11.62)
\end{aligned}
$$

where

$$F_1(y) = \frac{1}{N}\sum_{n=1,N}\delta(y - y_n), \qquad (11.63)$$

is the interference/diffraction function for N identical parallel slits of negligible width which are equally-spaced a distance d apart, and $F_1(y)$ is the corresponding aperture function. Furthermore [*cf.*, Equation (11.52)],

$$\bar{F}_2(\theta) = \int_{-\infty}^{\infty} F_2(y)\,\cos(k\,\sin\theta\,y)\,dy = \mathrm{sinc}[\pi\,(\delta/\lambda)\,\sin\theta], \qquad (11.64)$$

is the interference/diffraction function for a single slit of width δ.

We conclude, from the above analysis, that the interference/diffraction function for N identical equally-spaced parallel slits of finite width is the product of the interference/diffraction function for N identical equally-spaced parallel slits of negligible width, $\bar{F}_1(\theta)$, and the interference/diffraction function for a single slit of finite width, $\bar{F}_2(\theta)$. We have already encountered both of these functions. The former function (see Figure 11.8, which shows $[\bar{F}_1(\theta)]^2$) consists of a series of sharp maxima of equal amplitude located at [see Equation (11.40)]

$$\theta_j = \sin^{-1}\left(j\,\frac{\lambda}{d}\right), \qquad (11.65)$$

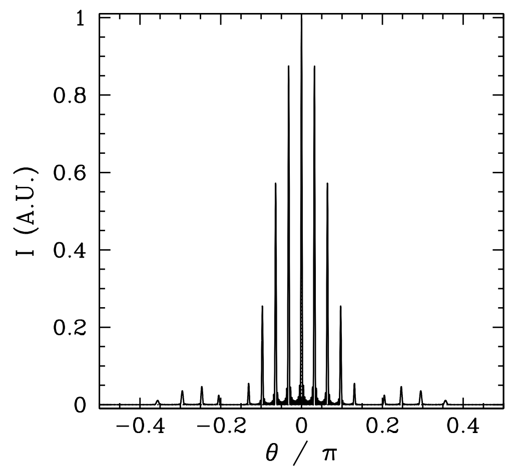


Figure 11.11: *Multi-slit far-field interference pattern calculated for* $N = 10$*,* $d/\lambda = 10$*, and* $\delta/\lambda = 2$*, assuming normal incidence.*

where j is an integer. The latter function (see Figure 11.9, which shows $[\bar{F}_2(\theta - \theta_0)]^2$) is of order unity for $|\theta| \lesssim \sin^{-1}(\lambda/\delta)$, and much less than unity for $|\theta| \gtrsim \sin^{-1}(\lambda/\delta)$. It follows that the intensity of the interference/diffraction pattern associated with N identical equally-spaced parallel slits of finite width, which is given by

$$\mathcal{I}(\theta) \propto \left[\bar{F}_1(\theta)\,\bar{F}_2(\theta)\right]^2 \propto \left[\bar{F}_1(\theta)\right]^2\,\left[\bar{F}_1(\theta)\right]^2, \tag{11.66}$$

is similar to that for N identical equally-spaced parallel slits of negligible width, $[\bar{F}_1(\theta)]^2$, except that the heights of the various maxima in the pattern are modulated by $[\bar{F}_2(\theta)]^2$. Hence, those maxima lying in the angular range $|\theta| < \sin^{-1}(\lambda/\delta)$ are of similar height, whereas those lying in the range $|\theta| > \sin^{-1}(\lambda/\delta)$ are of negligible height. This is illustrated in Figure 11.11, which shows the multi-slit interference/diffraction pattern calculated for $N = 10$, $d/\lambda = 20$, and $\delta/\lambda = 2$. As expected, the maxima lying in the angular range $|\theta| < \sin^{-1}(0.5) = \pi/6$ have relatively large amplitudes, whereas those lying in the range $|\theta| > \pi/6$ have negligibly small amplitudes.

11.8 Exercises

1. (a) Consider the geometric series

$$S = \sum_{n=0,N-1} z^n,$$

where z is a complex number. Demonstrate that

$$S = \frac{1-z^N}{1-z}.$$

(b) Suppose that $z = e^{\,i\,\theta}$, where θ is real. Employing the well-known identity

$$\sin\theta \equiv \frac{1}{2\,i}\left(e^{\,i\,\theta} - e^{-i\,\theta}\right),$$

show that

$$S = e^{\,i\,(N-1)\,\theta/2}\,\frac{\sin(N\,\theta/2)}{\sin(\theta/2)}.$$

(c) Finally, making use of de Moivre's theorem,

$$e^{\,i\,n\,\theta} \equiv \cos(n\,\theta) + i\,\sin(n\,\theta),$$

demonstrate that

$$C = \sum_{n=1,N} \cos(\alpha\,y_n),$$

where

$$y_n = [n-(N+1)/2]\,d,$$

evaluates to

$$C = \frac{\sin(N\,\alpha\,d/2)}{\sin(\alpha\,d/2)}.$$

2. An interference experiment employs two narrow parallel slits of separation 0.25 mm, and monochromatic light of wavelength $\lambda = 500\,\text{nm}$. Estimate the minimum distance that the projection screen must be placed behind the slits in order to obtain a far-field interference pattern.

3. A double-slit of slit separation 0.5 mm is illuminated at normal incidence by a parallel beam from a helium-neon laser that emits monochromatic light of wavelength 632.8 nm. A projection screen is located 5 m behind the slit. What is the separation of the central interference fringes on the screen?

4. Consider a double-slit interference experiment in which the slit spacing is 0.1 mm, and the projection screen is located 50 cm behind the slits. Assuming monochromatic illumination at normal incidence, if the observed separation between neighboring interference maxima at the center of the projection screen is 2.5 mm, what is the wavelength of the light illuminating the slits?

5. What is the mean length of the classical wavetrain (wave packet) corresponding to the light emitted by an atom whose excited state has a mean lifetime $\tau \sim 10^{-8}$ s? In an ordinary gas-discharge source, the excited atomic states do not decay freely, but instead have an effective lifetime $\tau \sim 10^{-9}$ s, due to collisions and Doppler effects. What is the length of the corresponding classical wavetrain?

6. If a “monochromatic” incoherent “line” source of visible light is not really a line, but has a finite width of 1 mm, estimate the minimum distance it can be placed in front of a double-slit, of slit separation 0.5 mm, if the light from the slit is to generate a clear interference pattern.

7. The visible emission spectrum of a sodium atom is dominated by a yellow line which actually consists of two closely-spaced lines of wavelength 589.0 nm and 589.6 nm. Demonstrate that a diffraction grating must have at least 328 lines in order to resolve this doublet at the third spectral order.

8. Consider a diffraction grating having 5000 lines per centimeter. Find the angular locations of the principle maxima when the grating is illuminated at normal incidence by (a) red light of wavelength 700 nm, and (b) violet light of wavelength 400 nm.

9. Suppose that a monochromatic laser of wavelength 632.8 nm emits a diffraction-limited beam of initial diameter 2 mm. Estimate how large a light spot the beam would produce on the surface of the Moon (which is a mean distance 3.76×10^5 km from the surface of the Earth). Neglect any effects of the Earth’s atmosphere.

10. Estimate how far away an automobile is when you can only just barely resolve the two headlights with your eyes.

11. Venus has a diameter of about 8000 miles. When it is prominently visible in the sky, in the early morning or late evening, it is about as far away as the Sun: *i.e.*, about 93 million miles. Now, Venus commonly appears larger than a point to the unaided eye. Are we seeing the true size of Venus?

12. The world’s largest steerable radio telescope, at the National Radio Astronomy Observatory, Green Bank, West Virginia, consists of a parabolic disk which is 300 ft in diameter. Estimate the angular resolution (in minutes of an arc) of the telescope when it is observing the well-known 21-cm radiation of hydrogen.

13. Estimate how large the lens of a camera carried by an artificial satellite orbiting the Earth at an altitude of 150 miles would have to be in order to resolve features on the Earth’s surface a foot in diameter.

14. Demonstrate that the secondary maxima in the far-field interference pattern generated by three identical equally-spaced parallel slits of negligible width are nine times less intense than the principle maxima.

15. Consider a double-slit interference/diffraction experiment in which the slit spacing is d, and the slit width δ. Show that the intensity of the far-field interference pattern, assuming normal incidence by monochromatic light of wavelength λ, is

$$\mathcal{I}(\theta) \propto \cos^2\left(\pi\,\frac{d}{\lambda}\,\sin\theta\right)\,\mathrm{sinc}^2\left(\pi\,\frac{\delta}{\lambda}\,\sin\theta\right).$$

Plot the intensity pattern for $d/\lambda = 8$ and $\delta/\lambda = 2$.

12 Wave Mechanics

12.1 Introduction

According to *classical physics* (*i.e.*, physics prior to the 20th century), *particles* and *waves* are two completely distinct classes of physical entity that possess markedly different properties. 1) Particles are *discrete*: *i.e.*, they cannot be arbitrarily divided. In other words, it makes sense to talk about one electron, or two electrons, but not about a third of an electron. Waves, on the other hand, are *continuous*: *i.e.*, they can be arbitrarily divided. In other words, given a wave whose amplitude has a certain value, it makes sense to talk about a similar wave whose amplitude is one third, or any other fraction whatsoever, of this value. 2) Particles are *highly localized* in space. For example, atomic nuclei have very small radii of order 10^{-15} m, whilst electrons act like point particles: *i.e.*, they have no discernible spatial extent. Waves, on the other hand, are *non-localized* in space. In fact, a wave is defined to be a disturbance that is periodic in space, with some finite periodicity length: *i.e.*, wavelength. Hence, it is fairly meaningless to talk about a disturbance being a wave unless it extends over a region of space that is at least a few wavelengths in dimension.

The classical scenario, described above, in which particles and waves are completely distinct from one another, had to be significantly modified in the early decades of the 20th century. During this time period, physicists discovered, much to their surprise, that, under certain circumstances, waves act as particles, and particles act as waves. This bizarre phenomenon is known as *wave-particle duality*. For instance, the *photoelectric effect* (see Section 12.2) indicates that electromagnetic waves sometimes act like swarms of massless particles called *photons*. Moreover, the phenomenon of *electron diffraction* by atomic lattices (see Section 12.3) implies that electrons sometimes have wave-like properties. Note, however, that wave-particle duality usually only manifests itself on atomic and sub-atomic lengthscales (*i.e.*, on lengthscales less than, or of order, 10^{-10} m—see Section 12.3). The classical picture remains valid on significantly longer lengthscales. In other words, on *macroscopic* lengthscales, waves only act like waves, particles only act like particles, and there is no wave-particle duality. However, on *microscopic* lengthscales, *classical mechanics*, which governs the macroscopic behavior of massive particles, and *classical electrodynamics*, which governs the macroscopic behavior of electromagnetic fields—neither of which take wave-particle duality into account—must be replaced by new theories. The theories in question are called *quantum mechanics* and *quantum electrodynamics*, respectively. In the following, we shall discuss a simplified version of quantum mechanics in which the microscopic dynamics of massive particles (*i.e.*, particles with finite mass) is described *entirely* in terms of wavefunctions. This particular theory is known as *wave mechanics*.

12.2 Photoelectric Effect

The so-called *photoelectric effect*, by which a polished metal surface emits electrons when illuminated by visible or ultra-violet light, was discovered by Heinrich Hertz in 1887. The following facts regarding this effect can be established via careful observation. First, a given surface only emits electrons when the *frequency* of the light with which it is illuminated exceeds a certain threshold value, which is a property of the metal. Second, the current of photoelectrons, when it exists, is proportional to the *intensity* of the light falling on the surface. Third, the energy of the photoelectrons is independent of the light intensity, but varies *linearly* with the light frequency. These facts are inexplicable within the framework of classical physics.

In 1905, Albert Einstein proposed a radical new theory of light in order to account for the photoelectric effect. According to this theory, light of fixed angular frequency ω consists of a collection of indivisible discrete packages, called *quanta*,[1] whose energy is

$$E = \hbar\,\omega. \tag{12.1}$$

Here, $\hbar = 1.055 \times 10^{-34}\,\mathrm{J\,s}$ is a new constant of nature, known as *Planck's constant*. (Strictly speaking, it is Planck's constant divided by 2π). Incidentally, $\hbar$ is called Planck's constant, rather than Einstein's constant, because Max Planck first introduced the concept of the quantization of light, in 1900, whilst trying to account for the electromagnetic spectrum of a black body (*i.e.*, a perfect emitter and absorber of electromagnetic radiation).

Suppose that the electrons at the surface of a piece of metal lie in a potential well of depth W. In other words, the electrons have to acquire an energy W in order to be emitted from the surface. Here, W is generally called the *work-function* of the surface, and is a property of the metal. Suppose that an electron absorbs a single quantum of light, otherwise known as a *photon*. Its energy therefore increases by $\hbar\,\omega$. If $\hbar\,\omega$ is greater than

[1] Plural of *quantum*: Latin neuter of *quantus*: how much.

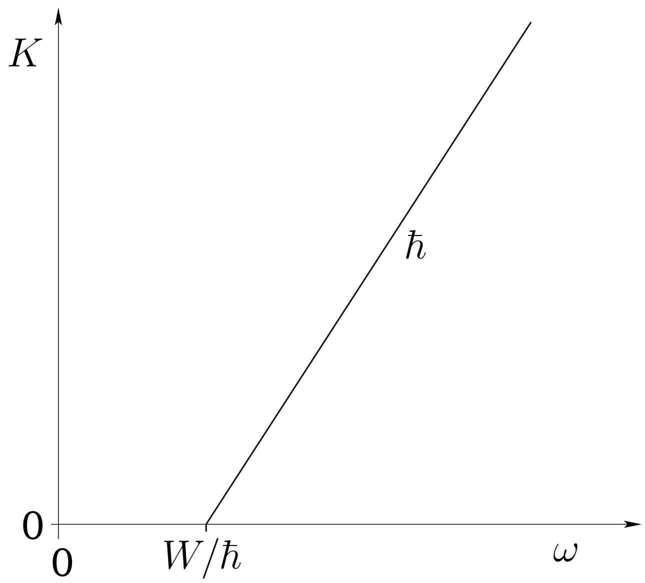


Figure 12.1: *Variation of the kinetic energy* K *of photoelectrons with the wave angular frequency* ω.

W then the electron is emitted from the surface with the residual kinetic energy

$$K = \hbar\,\omega - W. \tag{12.2}$$

Otherwise, the electron remains trapped in the potential well, and is not emitted. Here, we are assuming that the probability of an electron absorbing two or more photons is negligibly small compared to the probability of it absorbing a single photon (as is, indeed, the case for low intensity illumination). Incidentally, we can determine Planck's constant, as well as the work-function of the metal, by plotting the kinetic energy of the emitted photoelectrons as a function of the wave frequency, as shown in Figure 12.1. This plot is a straight-line whose slope is $\hbar$, and whose intercept with the ω axis is $W/\hbar$. Finally, the number of emitted electrons increases with the intensity of the light because the more intense the light the larger the flux of photons onto the surface. Thus, Einstein's quantum theory of light is capable of accounting for all three of the previously mentioned observational facts regarding the photoelectric effect. In the following, we shall assume that the central component of Einstein's theory—namely, Equation (12.1)—is a general result which applies to *all* particles, not just photons.

12.3 Electron Diffraction

In 1927, George Paget Thomson discovered that if a beam of electrons is made to pass through a thin metal film then the regular atomic array in the metal acts as a sort of diffraction grating, so that when a photographic film, placed behind the metal, is developed an *interference pattern* is discernible. Of course, this implies that electrons have wave-like properties. Moreover, the electron wavelength, λ, or, alternatively, the wavenumber, $k = 2\pi/\lambda$, can be deduced from the spacing of the maxima in the interference pattern (see Chapter 11). Thomson found that the momentum, p, of an electron is related to its wavenumber, k, according to the following simple relation:

$$p = \hbar\, k. \tag{12.3}$$

The associated wavelength, $\lambda = 2\pi/k$, is known as the *de Broglie wavelength*, since the above relation was first hypothesized by Louis de Broglie in 1926. In the following, we shall assume that Equation (12.3) is a general result which applies to *all* particles, not just electrons.

It turns out that wave-particle duality only manifests itself on length-scales less than, or of order, the de Broglie wavelength. Note, however, that this wavelength is generally pretty small. For instance, the de Broglie wavelength of an electron is

$$\lambda_e = 1.2 \times 10^{-9}\, [E(\mathrm{eV})]^{-1/2}\, \mathrm{m}, \tag{12.4}$$

where the electron energy is conveniently measured in units of electron-volts (eV). (An electron accelerated from rest through a potential difference of $1000\,\mathrm{V}$ acquires an energy of $1000\,\mathrm{eV}$, and so on. Electrons in atoms typically have energies in the range 10 to 100 eV.) Moreover, the de Broglie wavelength of a proton is

$$\lambda_p = 2.9 \times 10^{-11}\, [E(\mathrm{eV})]^{-1/2}\, \mathrm{m}. \tag{12.5}$$

12.4 Representation of Waves via Complex Numbers

In mathematics, the symbol i is conventionally used to represent the *square-root of minus one*: *i.e.*, the solution of $i^2 = -1$. Now, a *real number*, x (say), can take any value in a continuum of different values lying between $-\infty$ and $+\infty$. On the other hand, an *imaginary number* takes the general form $i\,y$, where y is a real number. It follows that the square of a real number is a positive real number, whereas the square of an imaginary number is a negative real number. In addition, a general *complex number* is written

$$z = x + i\,y, \tag{12.6}$$

where x and y are real numbers. In fact, x is termed the *real part* of z, and y the *imaginary part* of z. This is written mathematically as $x = \mathrm{Re}(z)$ and $y = \mathrm{Im}(z)$. Finally, the *complex conjugate* of z is defined $z^* = x - i\,y$.

Now, just as we can visualize a real number as a point on an infinite straight-line, we can visualize a complex number as a point in an infinite plane. The coordinates of the point in question are the real and imaginary parts of the number: *i.e.*, $z \equiv (x,\, y)$. This idea is illustrated in Figure 12.2. The distance, $r = \sqrt{x^2 + y^2}$, of the representative point from the origin is termed the *modulus* of the corresponding complex number, z. This is written mathematically as $|z| = \sqrt{x^2 + y^2}$. Incidentally, it follows that $z\,z^* = x^2 + y^2 = |z|^2$. The angle, $\theta = \tan^{-1}(y/x)$, that the straight-line joining the representative point to the origin subtends with the real axis is termed the *argument* of the corresponding complex number, z. This is written mathematically as $\arg(z) = \tan^{-1}(y/x)$. It follows from standard trigonometry that $x = r\,\cos\theta$, and $y = r\,\sin\theta$. Hence, $z = r\,\cos\theta + i\,r\sin\theta$.

Complex numbers are often used to represent waves, and wavefunctions. All such representations depend ultimately on a fundamental mathematical identity, known as *de Moivre's theorem* (see Exercise 12.1), which takes the form

$$e^{\,i\,\phi} \equiv \cos\phi + i\,\sin\phi, \tag{12.7}$$

where ϕ is a real number. Incidentally, given that $z = r\,\cos\theta + i\,r\,\sin\theta = r\,[\cos\theta + i\,\sin\theta]$, where z is a general complex number, $r = |z|$ its modulus, and $\theta = \arg(z)$ its argument, it follows from de Moivre's theorem that any complex number, z, can be written

$$z = r\,e^{\,i\,\theta}, \tag{12.8}$$

where $r = |z|$ and $\theta = \arg(z)$ are real numbers.

Now, a one-dimensional wavefunction takes the general form

$$\psi(x,t) = A\,\cos(\phi + k\,x - \omega\,t), \tag{12.9}$$

where $A > 0$ is the wave amplitude, ϕ the phase angle, k the wavenumber, and ω the angular frequency. Consider the complex wavefunction

$$\psi(x,t) = \psi_0\,e^{\,i\,(k\,x - \omega\,t)}, \tag{12.10}$$

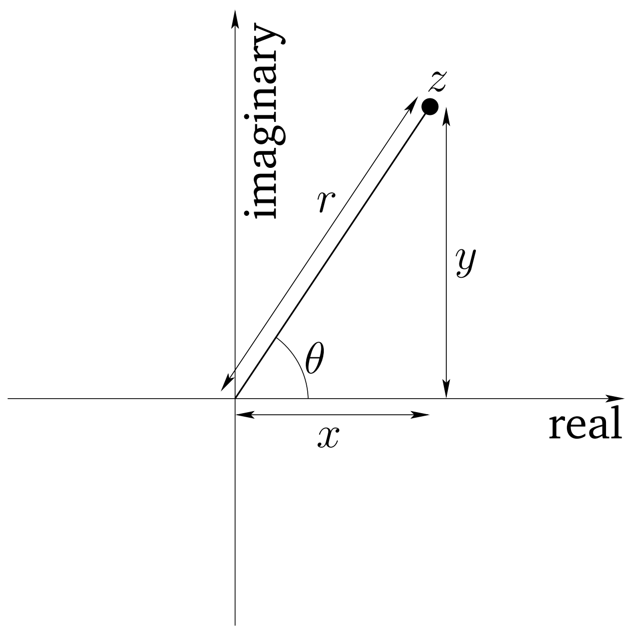


Figure 12.2: *Representation of a complex number as a point in a plane.*

where ψ_0 is a complex constant. We can write

$$\psi_0 = A\,e^{\,i\,\phi}, \tag{12.11}$$

where A is the modulus, and ϕ the argument, of ψ_0. Hence, we deduce that

$$\begin{aligned} \text{Re}\left[\psi_0\,e^{\,i\,(k\,x-\omega\,t)}\right] &= \text{Re}\left[A\,e^{\,i\,\phi}\,e^{\,i\,(k\,x-\omega\,t)}\right] \\ &= \text{Re}\left[A\,e^{\,i\,(\phi+k\,x-\omega\,t)}\right] \\ &= A\,\text{Re}\left[e^{\,i\,(\phi+k\,x-\omega\,t)}\right]. \end{aligned} \tag{12.12}$$

Thus, it follows from de Moirve's theorem, and Equation (12.9), that

$$\text{Re}\left[\psi_0\,e^{\,i\,(k\,x-\omega\,t)}\right] = A\,\cos(\phi + k\,x - \omega\,t) = \psi(x,t). \tag{12.13}$$

In other words, a general one-dimensional real wavefunction, (12.9), can be represented as the *real part* of a complex wavefunction of the form (12.10). For ease of notation, the "take the real part" aspect of the above expression is usually omitted, and our general one-dimension wavefunction is simply written

$$\psi(x,t) = \psi_0\,e^{\,i\,(k\,x-\omega\,t)}. \tag{12.14}$$

The main advantage of the complex representation, (12.14), over the more straightforward real representation, (12.9), is that the former enables us to combine the amplitude, A, and the phase angle, ϕ, of the wavefunction into a single complex amplitude, ψ_0.

12.5 Schrödinger's Equation

The basic premise of wave mechanics is that a massive particle of energy E and linear momentum p, moving in the x-direction (say), can be represented by a one-dimensional *complex wavefunction* of the form

$$\psi(x,t) = \psi_0\, e^{\,i\,(k\,x-\omega\,t)}, \tag{12.15}$$

where the complex amplitude, ψ_0, is arbitrary, whilst the wavenumber, k, and the angular frequency, ω, are related to the particle momentum, p, and energy, E, via the fundamental relations (12.3) and (12.1), respectively. Now, the above one-dimensional wavefunction is, presumably, the solution of some one-dimensional wave equation that determines how the wavefunction evolves in time. As described below, we can guess the form of this wave equation by drawing an analogy with classical physics.

A classical particle of mass m, moving in a one-dimensional potential $U(x)$, satisfies the energy conservation equation

$$E = K + U, \tag{12.16}$$

where

$$K = \frac{p^2}{2\,m} \tag{12.17}$$

is the particle's kinetic energy. Hence,

$$E\,\psi = (K+U)\,\psi \tag{12.18}$$

is a valid, but not obviously useful, wave equation.

However, it follows from Equations (12.15) and (12.1) that

$$\frac{\partial\psi}{\partial t} = -i\,\omega\,\psi_0\,e^{\,i\,(k\,x-\omega\,t)} = -i\,\frac{E}{\hbar}\,\psi, \tag{12.19}$$

which can be rearranged to give

$$E\,\psi = i\,\hbar\,\frac{\partial\psi}{\partial t}. \tag{12.20}$$

Likewise, from (12.15) and (12.3),

$$\frac{\partial^2\psi}{\partial x^2} = -k^2\,\psi_0\,e^{\,i\,(k\,x-\omega\,t)} = -\frac{p^2}{\hbar^2}\,\psi, \tag{12.21}$$

which can be rearranged to give

$$\frac{p^2}{2\,m}\,\psi = -\frac{\hbar^2}{2\,m}\,\frac{\partial^2\psi}{\partial x^2}. \tag{12.22}$$

Thus, combining Equations (12.18), (12.20), and (12.22), we obtain

$$i\,\hbar\,\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2\,m}\,\frac{\partial^2\psi}{\partial x^2} + U(x)\,\psi. \tag{12.23}$$

This equation, which is known as *Schrödinger's equation*—since it was first formulated by Erwin Schrödinder in 1926—is the fundamental equation of wave mechanics.

Now, for a massive particle moving in free space (*i.e.*, $U = 0$), the complex wavefunction (12.15) is a solution of Schrödinger's equation, (12.23), provided that

$$\omega = \frac{\hbar}{2\,m}\,k^2. \tag{12.24}$$

The above expression can be thought of as the *dispersion relation* (see Section 5.1) for matter waves in free space. The associated *phase velocity* (see Section 7.2) is

$$v_p = \frac{\omega}{k} = \frac{\hbar\,k}{2\,m} = \frac{p}{2\,m}, \tag{12.25}$$

where use has been made of (12.3). Note that this phase velocity is only *half* the classical velocity, $v = p/m$, of a massive (non-relativistic) particle.

12.6 Probability Interpretation of the Wavefunction

After many false starts, physicists in the early 20th century came to the conclusion that the only self-consistent physical interpretation of a particle wavefunction, which is consistent with experimental observations, is *probabilistic* in nature. To be more exact, if $\psi(x,t)$ is the complex wavefunction of a given particle, moving in one-dimension along the x-axis, then the *probability* of finding the particle between x and $x + dx$ at time t is

$$P(x,t) = |\psi(x,t)|^2\,dx. \tag{12.26}$$

A probability is, of course, a real number lying in the range 0 to 1. An event which has a probability 0 is impossible. On the other hand, an event which has a probability 1 is certain to occur. An event which has an probability 1/2 (say) is such that in a very large number of identical trials the event occurs in half of the trials. Now, we can interpret

$$P(t) = \int_{-\infty}^{\infty} |\psi(x,t)|^2 \, dx \tag{12.27}$$

as the probability of the particle being found *anywhere* between $x = -\infty$ and $x = +\infty$ at time t. This follows, via induction, from the fundamental result in probability theory that the probability of the occurrence of *one or other* of two *mutually exclusive* events (such as the particle being found in two non-overlapping regions) is the *sum* (or integral) of the probabilities of the individual events. (For example, the probability of throwing a 1 on a six-sided die is 1/6. Likewise, the probability of throwing a 2 is 1/6. Hence, the probability of throwing a 1 *or* a 2 is $1/6 + 1/6 = 1/3$.) Now, assuming that the particle exists, it is *certain* that it will be found somewhere between $x = -\infty$ and $x = +\infty$ at time t. Since a certain event has probability 1, our probability interpretation of the wavefunction is only tenable provided that

$$\int_{-\infty}^{\infty} |\psi(x,t)|^2 \, dx = 1 \tag{12.28}$$

at all times. A wavefunction which satisfies the above condition is said to be *properly normalized*.

Suppose that we have a wavefunction, $\psi(x,t)$, which is such that it satisfies the normalization condition (12.28) at time $t = 0$. Furthermore, let the wavefunction evolve in time according to Schrödinger's equation, (12.23). Our probability interpretation of the wavefunction only makes sense if the normalization condition remains satisfied at all subsequent times. This follows because if the particle is certain to be found somewhere on the x-axis (which is the interpretation put on the normalization condition) at time $t = 0$ then it is equally certain to be found somewhere on the x-axis at a later time (since we are not presently dealing with any physical process by which particles can be created or destroyed). Thus, it is necessary for us to demonstrate that Schrödinger's equation preserves the normalization of the wavefunction.

Taking Schrödinger's equation, and multiplying it by ψ^* (the complex conjugate of the wavefunction), we obtain

$$i\,\hbar\,\frac{\partial\psi}{\partial t}\,\psi^* = -\frac{\hbar^2}{2\,m}\,\frac{\partial^2\psi}{\partial x^2}\,\psi^* + U(x)\,|\psi|^2. \tag{12.29}$$

The complex conjugate of the above expression yields

$$-\,i\,\hbar\,\frac{\partial\psi^*}{\partial t}\,\psi = -\frac{\hbar^2}{2\,m}\,\frac{\partial^2\psi^*}{\partial x^2}\,\psi + U(x)\,|\psi|^2. \tag{12.30}$$

Here, we have made use of the easily demonstrated results $(\psi^*)^* = \psi$ and $i^* = -i$, as well as the fact that U is real. Taking the difference between the above two expressions, we obtain

$$i\,\hbar\left(\frac{\partial\psi}{\partial t}\,\psi^* + \frac{\partial\psi^*}{\partial t}\,\psi\right) = -\frac{\hbar^2}{2\,m}\left(\frac{\partial^2\psi}{\partial x^2}\,\psi^* - \frac{\partial^2\psi^*}{\partial x^2}\,\psi\right), \tag{12.31}$$

which can be written

$$i\,\hbar\,\frac{\partial|\psi|^2}{\partial t} = -\frac{\hbar^2}{2\,m}\,\frac{\partial}{\partial x}\left(\frac{\partial\psi}{\partial x}\,\psi^* - \frac{\partial\psi^*}{\partial x}\,\psi\right). \tag{12.32}$$

Integrating in x, we get

$$i\,\hbar\,\frac{d}{dt}\int_{-\infty}^{\infty}|\psi|^2\,dx = -\frac{\hbar^2}{2\,m}\left[\frac{\partial\psi}{\partial x}\,\psi^* - \frac{\partial\psi^*}{\partial x}\,\psi\right]_{-\infty}^{\infty}. \tag{12.33}$$

Finally, assuming that the wavefunction is *localized* in space: *i.e.*,

$$|\psi(x,t)| \to 0 \quad \text{as} \quad |x| \to \infty, \tag{12.34}$$

we obtain

$$\frac{d}{dt}\int_{-\infty}^{\infty}|\psi|^2\,dx = 0. \tag{12.35}$$

It follows, from the above analysis, that if a localized wavefunction is properly normalized at $t = 0$ (*i.e.*, if $\int_{-\infty}^{\infty}|\psi(x,0)|^2\,dx = 1$) then it will remain properly normalized as it evolves in time according to Schrödinger's equation. Incidentally, a wavefunction which is not localized cannot be properly normalized, since its normalization integral $\int_{-\infty}^{\infty}|\psi|^2\,dx$ is necessarily *infinite*. For such a wavefunction, $|\psi(x,t)|^2\,dx$ gives the *relative probability*, rather than the absolute probability, of finding the particle between x and $x + dx$ at time t: *i.e.*, [*cf.*, Equation (12.26)]

$$P(x,t) \propto |\psi(x,t)|^2\,dx. \tag{12.36}$$

12.7 Wave Packets

As we have seen, the wavefunction of a massive particle of momentum p and energy E, moving in free space along the x-axis, can be written

$$\psi(x,t) = \bar{\psi}\, e^{\,i\,(k\,x-\omega\,t)}, \tag{12.37}$$

where $k = p/\hbar$, $\omega = E/\hbar$, and $\bar{\psi}$ is a complex constant. Here, ω and k are linked via the matter wave dispersion relation (12.24). Expression (12.37) represents a *plane wave* which propagates in the x-direction with the phase velocity $v_p = \omega/k$. However, according to (12.25), this phase velocity is only half of the classical velocity of a massive particle.

Now, according to the discussion in the previous section, the most reasonable physical interpretation of the wavefunction is that $|\psi(x,t)|^2\,dx$ is proportional to (assuming that the wavefunction is not properly normalized) the *probability* of finding the particle between x and $x + dx$ at time t. However, the modulus squared of the wavefunction (12.37) is $|\bar{\psi}|^2$, which is a constant that depends on neither x nor t. In other words, the above wavefunction represents a particle which is *equally likely* to be found *anywhere* on the x-axis *at all times*. Hence, the fact that this wavefunction propagates at a phase velocity which does not correspond to the classical particle velocity does not have any observable consequences.

So, how can we write the wavefunction of a particle which is *localized* in x: *i.e.*, a particle which is more likely to be found at some positions on the x-axis than at others? It turns out that we can achieve this goal by forming a *linear combination* of plane waves of different wavenumbers: *i.e.*,

$$\psi(x,t) = \int_{-\infty}^{\infty} \bar{\psi}(k)\, e^{\,i\,(k\,x-\omega\,t)}\, dk. \tag{12.38}$$

Here, $\bar{\psi}(k)$ represents the complex amplitude of plane waves of wavenumber k within this combination. In writing the above expression, we are relying on the assumption that matter waves are *superposable*: *i.e.*, it is possible to add two valid wave solutions to form a third valid wave solution. The ultimate justification for this assumption is that matter waves satisfy the *linear* wave equation (12.23).

Now, there is a fundamental mathematical theorem, known as *Fourier's theorem* (see Section 8.1 and Exercise 12.2), which states that if

$$f(x) = \int_{-\infty}^{\infty} \bar{f}(k)\, e^{\,i\,k\,x}\, dk, \tag{12.39}$$

then

$$\bar{f}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\,e^{-i\,k\,x}\,dx. \tag{12.40}$$

Here, $\bar{f}(k)$ is known as the *Fourier transform* of the function $f(x)$. We can use Fourier's theorem to find the k-space function $\bar{\psi}(k)$ which generates any given x-space wavefunction $\psi(x)$ at a given time.

For instance, suppose that at $t = 0$ the wavefunction of our particle takes the form

$$\psi(x,0) \propto \exp\left[i\,k_0\,x - \frac{(x-x_0)^{\,2}}{4\,(\Delta x)^{\,2}}\right]. \tag{12.41}$$

Thus, the initial probability distribution for the particle's x-coordinate is

$$|\psi(x,0)|^{\,2} \propto \exp\left[-\frac{(x-x_0)^{\,2}}{2\,(\Delta x)^{\,2}}\right]. \tag{12.42}$$

This particular distribution is called a *Gaussian* distribution (see Section 8.1), and is plotted in Figure 12.3. It can be seen that a measurement of the particle's position is most likely to yield the value x_0, and very unlikely to yield a value which differs from x_0 by more than $3\,\Delta x$. Thus, (12.41) is the wavefunction of a particle which is initially localized in some region of x-space, centered on $x = x_0$, whose width is of order Δx. This type of wavefunction is known as a *wave packet*. Of course, a wave packet is just another name for a wave pulse (see Chapter 8).

Now, according to Equation (12.38),

$$\psi(x,0) = \int_{-\infty}^{\infty} \bar{\psi}(k)\,e^{i\,k\,x}\,dk. \tag{12.43}$$

Hence, we can employ Fourier's theorem to invert this expression to give

$$\bar{\psi}(k) \propto \int_{-\infty}^{\infty} \psi(x,0)\,e^{-i\,k\,x}\,dx. \tag{12.44}$$

Making use of Equation (12.41), we obtain

$$\bar{\psi}(k) \propto e^{-i\,(k-k_0)\,x_0}\int_{-\infty}^{\infty} \exp\left[-i\,(k-k_0)\,(x-x_0) - \frac{(x-x_0)^2}{4\,(\Delta x)^2}\right] dx. \tag{12.45}$$

Changing the variable of integration to $y = (x-x_0)/(2\,\Delta x)$, the above expression reduces to

$$\bar{\psi}(k) \propto e^{-i\,k\,x_0-\beta^2/4}\int_{-\infty}^{\infty} e^{-(y-y_0)^{\,2}}\,dy, \tag{12.46}$$

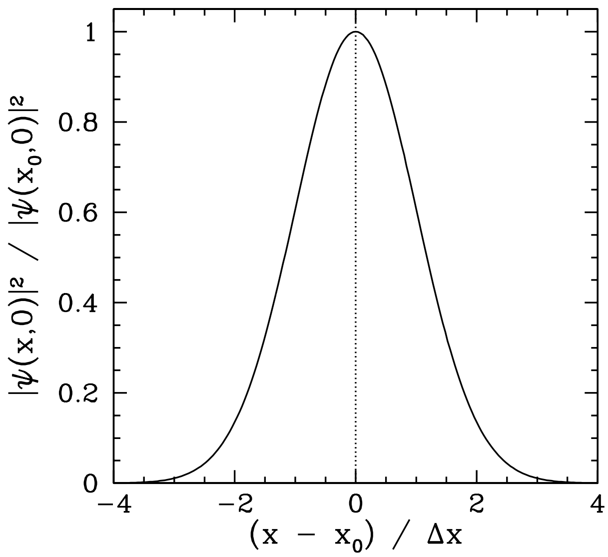


Figure 12.3: *A one-dimensional Gaussian probability distribution.*

where $\beta = 2\,(k-k_0)\,\Delta x$ and $y_0 = -i\,\beta/2$. The integral in the above equation is now just a number, as can easily be seen by making the second change of variable $z = y - y_0$. Hence, we deduce that

$$\bar{\psi}(k) \propto \exp\left[-i\,k\,x_0 - \frac{(k-k_0)^2}{4\,(\Delta k)^2}\right], \tag{12.47}$$

where

$$\Delta k = \frac{1}{2\,\Delta x}. \tag{12.48}$$

Now, if $|\psi(x,0)|^2\,dx$ is proportional to the probability of a measurement of the particle's position yielding a value in the range x to $x+dx$ at time $t = 0$ then it stands to reason that $|\bar{\psi}(k)|^2\,dk$ is proportional to the probability of a measurement of the particle's wavenumber yielding a value in the range k to $k + dk$. (Recall that $p = \hbar\,k$, so a measurement of the particle's wavenumber, k, is equivalent to a measurement of the particle's momentum, p). According to Equation (12.47),

$$|\bar{\psi}(k)|^2 \propto \exp\left[-\frac{(k-k_0)^2}{2\,(\Delta k)^2}\right]. \tag{12.49}$$

Note that this probability distribution is a *Gaussian* in k-space—see Equation (12.42) and Figure 12.3. Hence, a measurement of k is most likely to yield the value k_0, and very unlikely to yield a value which differs from k_0 by more than $3\,\Delta k$. Incidentally, as was previously mentioned in Section 8.1, a Gaussian is the *only* mathematical function in x-space which has the same form as its Fourier transform in k-space.

We have just seen that a wave packet with a Gaussian probability distribution of characteristic width Δx in x-space [see Equation (12.42)] is equivalent to a wave packet with a Gaussian probability distribution of characteristic width Δk in k-space [see Equation (12.49)], where

$$\Delta x\,\Delta k = \frac{1}{2}. \tag{12.50}$$

This illustrates an important property of wave packets. Namely, in order to construct a packet which is highly localized in x-space (*i.e.*, with small Δx) we need to combine plane waves with a very wide range of different k-values (*i.e.*, with large Δk). Conversely, if we only combine plane waves whose wavenumbers differ by a small amount (*i.e.*, if Δk is small) then the resulting wave packet is highly extended in x-space (*i.e.*, Δx is large).

Now, according to Section 9.1, a wave packet made up of a superposition of plane waves that is strongly peaked around some central wavenumber k_0 propagates at the *group velocity*,

$$v_g = \frac{d\omega(k_0)}{dk}, \tag{12.51}$$

rather than the *phase velocity*, $v_p = (\omega/k)_{k_0}$, assuming that all of the constituent plane waves satisfy a dispersion relation of the form $\omega = \omega(k)$. Now, for the case of matter waves, the dispersion relation is (12.24). Thus, the associated group velocity is

$$v_g = \frac{\hbar\,k_0}{m} = \frac{p}{m}, \tag{12.52}$$

where $p = \hbar\,k_0$. Note that this velocity is *identical* to the classical velocity of a (non-relativistic) massive particle. We conclude that the matter wave dispersion relation (12.24) is perfectly consistent with classical physics, as long as we recognize that particles must be identified with *wave packets* (which propagate at the group velocity) rather than plane waves (which propagate at the phase velocity).

In Section 9.1, it was also demonstrated that the spatial extent of a wave packet of initial extent $(\Delta x)_0$ grows, as the packet evolves in time, like

$$\Delta x \simeq (\Delta x)_0 + \frac{d^2\omega(k_0)}{dk^2}\,\frac{t}{(\Delta x)_0}, \tag{12.53}$$

where k_0 is the packet's central wavenumber. Thus, it follows from the matter wave dispersion relation, (12.24), that the width of a particle wave packet grows in time as

$$\Delta x \simeq (\Delta x)_0 + \frac{\hbar}{m}\,\frac{t}{(\Delta x)_0}. \tag{12.54}$$

For example, if an electron wave packet is initially localized in a region of atomic dimensions (*i.e.*, $\Delta x \sim 10^{-10}\,\mathrm{m}$) then the width of the packet doubles in about $10^{-16}\,\mathrm{s}$. Clearly, particle wave packets spread out very rapidly indeed (in free space).

12.8 Heisenberg's Uncertainty Principle

According to the analysis contained in the previous section, a particle wave packet that is initially localized in x-space, with characteristic width Δx, is also localized in k-space, with characteristic width $\Delta k = 1/(2\,\Delta x)$. However, as time progresses, the width of the wave packet in x-space increases [see Equation (12.54)], whilst that of the packet in k-space stays the same [since $\bar{\psi}(k)$ is given by Equation (12.44) at all times.] Hence, in general, we can say that

$$\Delta x\,\Delta k \gtrsim \frac{1}{2}. \tag{12.55}$$

Furthermore, we can interpret Δx and Δk as characterizing our *uncertainty* regarding the values of the particle's position and wavenumber, respectively.

Now, a measurement of a particle's wavenumber, k, is equivalent to a measurement of its momentum, p, since $p = \hbar\,k$. Hence, an uncertainty in k of order Δk translates to an uncertainty in p of order $\Delta p = \hbar\,\Delta k$. It follows, from the above inequality, that

$$\Delta x\,\Delta p \gtrsim \frac{\hbar}{2}. \tag{12.56}$$

This is the famous *Heisenberg uncertainty principle*, first proposed by Werner Heisenberg in 1927. According to this principle, it is impossible to *simultaneously* measure the position and momentum of a particle (exactly). Indeed, a good knowledge of the particle's position implies a poor knowledge of its momentum, and *vice versa*. Note that the uncertainty principle is a direct consequence of representing particles as waves.

It is apparent, from expression (12.54), that a particle wave packet of initial spatial extent $(\Delta x)_0$ spreads out in such a manner that its spatial extent becomes

$$\Delta x \sim \frac{\hbar\, t}{m\,(\Delta x)_0} \tag{12.57}$$

at large t. It is easily demonstrated that this spreading of the wave packet is a consequence of the uncertainty principle. Indeed, since the initial uncertainty in the particle's position is $(\Delta x)_0$, it follows that the uncertainty in its momentum is of order $\hbar/(\Delta x)_0$. This translates to an uncertainty in velocity of $\Delta v = \hbar/[m\,(\Delta x)_0]$. Thus, if we imagine that part of the wave packet propagates at $v_0 + \Delta v/2$, and another part at $v_0 - \Delta v/2$, where v_0 is the mean propagation velocity, then it is clear that the wave packet will spread out as time progresses. Indeed, at large t, we expect the width of the wave packet to be

$$\Delta x \sim \Delta v\, t \sim \frac{\hbar\, t}{m\,(\Delta x)_0}, \tag{12.58}$$

which is identical to Equation (12.57). Evidently, the spreading of a particle wave packet, as time progresses, should be interpreted as representing an increase in our *uncertainty* regarding the particle's position, rather than an increase in the spatial extent of the particle itself.

12.9 Collapse of the Wavefunction

Consider a spatially extended wavefunction, $\psi(x,t)$. According to our usual interpretation, $|\psi(x,t)|^2\,dx$ is proportional to the probability of a measurement of the particle's position yielding a value in the range x to $x + dx$ at time t. Thus, if the wavefunction is extended then there is a wide range of likely values that such a measurement could give. Suppose, however, that we make a measurement of the particle's position, and obtain the value x_0. We now know that the particle is located at $x = x_0$. If we make another measurement, immediately after the first one, then what value would we expect to obtain? Well, common sense tells us that we should obtain the *same* value, x_0, since the particle cannot have shifted position appreciably in an infinitesimal time interval. Thus, immediately after the first measurement, a measurement of the particle's position is *certain* to give the value x_0, and has no chance of giving any other value. This implies that the wavefunction must have *collapsed* to some sort of "spike" function, centered on $x = x_0$. This idea is illustrated in Figure 12.4. Of course, as soon as the wavefunction collapses, it starts to expand again, as described in the previous section. Thus, the second measurement must be made reasonably quickly after the first one, otherwise the same result will not necessarily be obtained.

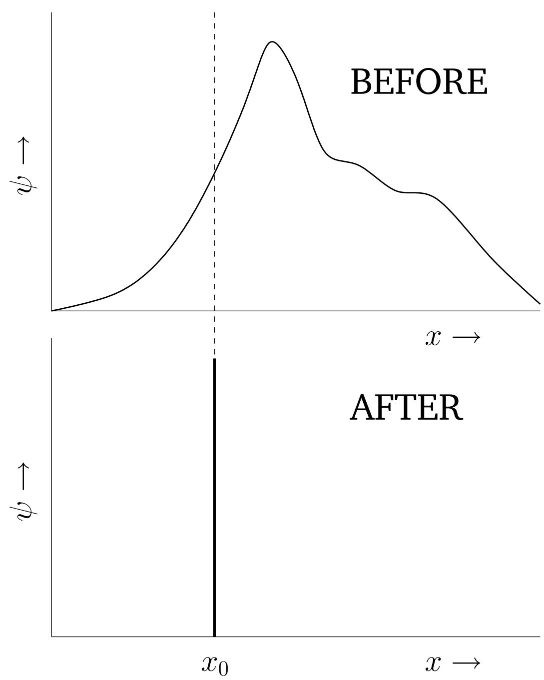


Figure 12.4: *Collapse of the wavefunction upon measurement of* x*.*

The above discussion illustrates an important point in wave mechanics. Namely, that the wavefunction of a massive particle changes *discontinuously* (in time) whenever a measurement of the particle's position is made. We conclude that there are two types of time evolution of the wavefunction in wave mechanics. First, there is a *smooth* evolution which is governed by Schrödinger's equation. This evolution takes place *between* measurements. Second, there is a *discontinuous* evolution which takes place each time a measurement is made.

12.10 Stationary States

Consider *separable* solutions to Schrödinger's equation of the form

$$\psi(x,t) = \psi(x)\, e^{-i\,\omega\,t}. \tag{12.59}$$

According to (12.20), such solutions have definite energies $E = \hbar\,\omega$. For this reason, they are usually written

$$\psi(x,t) = \psi(x)\, e^{-i\,(E/\hbar)\,t}. \tag{12.60}$$

Now, the probability of finding the particle between x and $x + dx$ at time t is

$$P(x,t) = |\psi(x,t)|^2\, dx = |\psi(x)|^2\, dx. \tag{12.61}$$

Note that this probability is *time independent*. For this reason, wavefunctions of the form (12.60) are known as *stationary states*. Moreover, $\psi(x)$ is called a *stationary wavefunction*. Substituting (12.60) into Schrödinger's equation, (12.23), we obtain the following expression for $\psi(x)$:

$$-\frac{\hbar^2}{2\,m}\,\frac{d^2\psi}{dx^2} + U(x)\,\psi = E\,\psi. \tag{12.62}$$

Not surprisingly, the above equation is called the *time independent Schrödinger equation*.

Consider a particle trapped in a one-dimensional square potential well, of infinite depth, which is such that

$$U(x) = \left\{ \begin{array}{lll} 0 & & 0 \leq x \leq a \\ \infty & & \text{otherwise} \end{array} \right. . \tag{12.63}$$

The particle is obviously excluded from the region $x < 0$ or $x > a$, so $\psi = 0$ in this region (*i.e.*, there is zero probability of finding the particle outside the well). Within the well, a particle of definite energy E has a stationary wavefunction, $\psi(x)$, which satisfies

$$-\frac{\hbar^2}{2\,m}\,\frac{d^2\psi}{dx^2} = E\,\psi. \tag{12.64}$$

The boundary conditions are

$$\psi(0) = \psi(a) = 0. \tag{12.65}$$

This follows because $\psi = 0$ in the region $x < 0$ or $x > a$, and $\psi(x)$ must be *continuous* [since a discontinuous wavefunction would generate a singular term (*i.e.*, the term involving $d^2\psi/dx^2$) in the time independent Schrödinger equation, (12.62), which could not be balanced, even by an infinite potential].

Let us search for solutions to (12.64) of the form

$$\psi(x) = \psi_0\,\sin(k\,x), \tag{12.66}$$

where ψ_0 is a constant. It follows that

$$\frac{\hbar^2\,k^2}{2\,m} = E. \tag{12.67}$$

The solution (12.66) automatically satisfies the boundary condition $\psi(0) = 0$. The second boundary condition, $\psi(a) = 0$, leads to a quantization of the wavenumber: *i.e.*,

$$k = n\,\frac{\pi}{a}, \tag{12.68}$$

where $n = 1, 2, 3,$ *etc.* (Note that a "quantized" quantity is one which can only take discrete values.) According to (12.67), the energy is also quantized. In fact, $E = E_n$, where

$$E_n = n^2\,\frac{\hbar^2\,\pi^2}{2\,m\,a^2}. \tag{12.69}$$

Thus the allowed wavefunctions for a particle trapped in a one-dimensional square potential well of infinite depth are

$$\psi_n(x,t) = A_n\,\sin\left(n\,\pi\,\frac{x}{a}\right)\,\exp\left(-i\,n^2\,\frac{E_1}{\hbar}\,t\right), \tag{12.70}$$

where n is a positive integer, and A_n a constant. Note that we cannot have $n = 0$, since, in this case, we obtain a null wavefunction: *i.e.*, $\psi = 0$, everywhere. Furthermore, if n takes a negative integer value then it generates exactly the same wavefunction as the corresponding positive integer value (assuming $\psi_{-n} = -\psi_n$).

The constant A_n, appearing in the above wavefunction, can be determined from the constraint that the wavefunction be properly normalized. For the problem presently under consideration, the normalization condition (12.28) reduces to

$$\int_0^a |\psi(x)|^2\,dx = 1. \tag{12.71}$$

It follows from (12.70) that $|A_n|^2 = 2/a$. Hence, a properly normalized version of the wavefunction (12.70) is

$$\psi_n(x,t) = \left(\frac{2}{a}\right)^{1/2} \sin\left(n\,\pi\,\frac{x}{a}\right)\,\exp\left(-i\,n^2\,\frac{E_1}{\hbar}\,t\right). \tag{12.72}$$

Figure 12.5 shows the first four properly normalized stationary wavefunctions for a particle trapped in a one-dimensional square potential well of infinite depth: *i.e.*, $\psi_n(x) = \sqrt{2/a}\,\sin(n\,\pi\,x/a)$, for $n = 1$ to 4.

Note that the stationary wavefunctions that we have just found are, in essence, *standing wave* solutions to Schrödinger's equation. Indeed, the wavefunctions are very similar in form to the classical standing wave solutions discussed in Chapters 5 and 6.

At first sight, it seems rather strange that the lowest energy that a particle trapped in a one-dimensional potential well can have is not zero, as would be the case in classical mechanics, but rather $E_1 = \hbar^2\,\pi^2/(2\,m\,a^2)$. In fact, as explained in the following, this residual energy is a direct consequence of *Heisenberg's uncertainty principle*. Now, a particle trapped in a one-dimensional well of width a is likely to be found anywhere inside the well. Thus, the uncertainty in the particle's position is $\Delta x \sim a$. It follows from the uncertainty principle, (12.56), that

$$\Delta p \gtrsim \frac{\hbar}{2\,\Delta x} \sim \frac{\hbar}{a}. \tag{12.73}$$

In other words, the particle *cannot* have zero momentum. In fact, the particle's momentum must be at least $p \sim \hbar/a$. However, for a free particle, $E = p^2/2\,m$. Hence, the residual energy associated with the particle's residual momentum is

$$E \sim \frac{p^2}{m} \sim \frac{\hbar^2}{m\,a^2} \sim E_1. \tag{12.74}$$

This type of residual energy, which is often found in quantum mechanical systems, and has no equivalent in classical mechanics, is generally known as *zero point energy*.

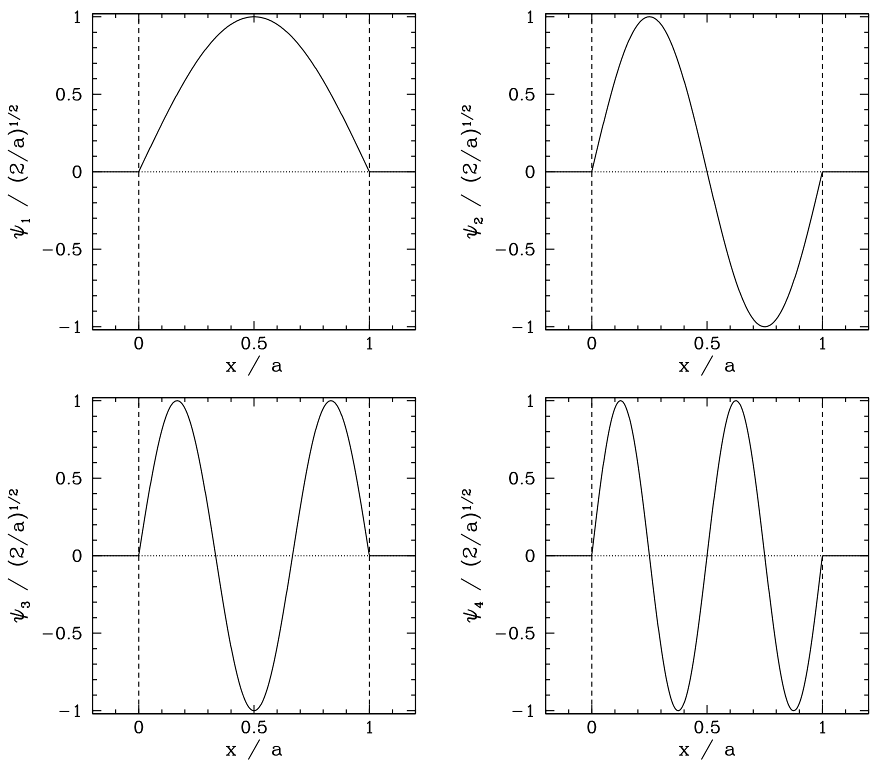


Figure 12.5: *First four stationary wavefunctions for a particle trapped in a one-dimensional square potential well of infinite depth.*

12.11 Three-Dimensional Wave Mechanics

Up to now, we have only discussed wave mechanics for a particle moving in one dimension. However, the generalization to a particle moving in three dimensions is fairly straightforward. A massive particle moving in three dimensions has a complex wavefunction of the form [*cf.*, (12.15)]

$$\psi(x, y, z, t) = \psi_0\, e^{\,i\,(\mathbf{k}\cdot\mathbf{r}-\omega\,t)}, \tag{12.75}$$

where ψ_0 is a complex constant, and $\mathbf{r} = (x,\, y,\, z)$. Here, the wavevector, $\mathbf{k}$, and the angular frequency, ω, are related to the particle momentum, $\mathbf{p}$, and energy, E, according to [*cf.*, (12.3)]

$$\mathbf{p} = \hbar\,\mathbf{k}, \tag{12.76}$$

and [*cf.*, (12.1)]

$$E = \hbar\,\omega, \tag{12.77}$$

respectively. Generalizing the analysis of Section 12.5, the three-dimensional version of Schrödinger's equation is easily shown to take the form [*cf.*, (12.23)]

$$i\,\hbar\,\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2\,m}\,\nabla^2\psi + U(\mathbf{r})\,\psi, \tag{12.78}$$

where the differential operator

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{12.79}$$

is known as the *Laplacian*. The interpretation of a three-dimensional wavefunction is that the probability of finding the particle between x and $x + dx$, between y and $y + dy$, and between z and $z + dz$, at time t is [*cf.*, (12.26)]

$$P(x, y, z, t) = |\psi(x, y, z, t)|^2\, dx\, dy\, dz. \tag{12.80}$$

Moreover, the normalization condition for the wavefunction becomes [*cf.*, (12.28)]

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\psi(x, y, z, t)|^2\, dx\, dy\, dz = 1. \tag{12.81}$$

Incidentally, it is easily demonstrated that Schrödinger's equation, (12.78), preserves the normalization condition, (12.81), of a *localized* wavefunction.

Heisenberg's uncertainty principle generalizes to [*cf.*, (12.56)]

$$\Delta x\,\Delta p_x \gtrsim \frac{\hbar}{2}, \tag{12.82}$$

$$\Delta y\,\Delta p_y \gtrsim \frac{\hbar}{2}, \tag{12.83}$$

$$\Delta z\,\Delta p_z \gtrsim \frac{\hbar}{2}. \tag{12.84}$$

Finally, a stationary state of energy E is written [*cf.*, (12.60)]

$$\psi(x,y,z,t) = \psi(x,y,z)\,e^{-i\,(E/\hbar)\,t}, \tag{12.85}$$

where the stationary wavefunction, $\psi(x,y,z)$, satisfies [*cf.*, (12.62)]

$$-\frac{\hbar^2}{2\,m}\,\nabla^2\psi + U(\mathbf{r})\,\psi = E\,\psi. \tag{12.86}$$

As an example of a three-dimensional problem in wave mechanics, consider a particle trapped in a square potential well of infinite depth which is such that

$$U(x,y,z) = \begin{cases} 0 & 0 \le x \le a,\ 0 \le y \le a,\ 0 \le z \le a \\ \infty & \text{otherwise} \end{cases}. \tag{12.87}$$

Within the well, the stationary wavefunction, $\psi(x,y,z)$, satisfies

$$-\frac{\hbar^2}{2\,m}\,\nabla^2\psi = E\,\psi, \tag{12.88}$$

subject to the boundary conditions

$$\psi(0,y,z) = \psi(x,0,z) = \psi(x,y,0) = 0, \tag{12.89}$$

and

$$\psi(a,y,z) = \psi(x,a,z) = \psi(x,y,a) = 0, \tag{12.90}$$

since $\psi = 0$ outside the well. Let us try a seperable wavefunction of the form

$$\psi(x,y,z) = \psi_0\,\sin(k_x\,x)\,\sin(k_y\,y)\,\sin(k_z\,z). \tag{12.91}$$

This expression automatically satisfies the boundary conditions (12.89). The remaining boundary conditions, (12.90), are satisfied provided

$$k_x = n_x\,\frac{\pi}{a}, \tag{12.92}$$

$$k_y = n_y\,\frac{\pi}{a}, \tag{12.93}$$

$$k_z = n_z\,\frac{\pi}{a}, \tag{12.94}$$

where n_x, n_y, and n_z are (independent) *positive integers*. Substitution of the wavefunction (12.91) into Equation (12.88) yields

$$E = \frac{\hbar^2}{2\,m}\,(k_x^{\,2}+k_y^{\,2}+k_z^{\,2}). \tag{12.95}$$

Thus, it follows from Equations (12.92)–(12.94) that the particle energy is quantized, and that the allowed *energy levels* are

$$E_{n_x,n_y,n_z} = \frac{\hbar^2}{2\,m\,a^2}\,(n_x^{\,2}+n_y^{\,2}+n_z^{\,2}). \tag{12.96}$$

The properly normalized [see Equation (12.81)] stationary wavefunctions corresponding to these energy levels are

$$\psi_{n_x,n_y,n_z}(x,y,z) = \left(\frac{2}{a}\right)^{3/2}\,\sin\left(n_x\,\pi\,\frac{x}{a}\right)\,\sin\left(n_y\,\pi\,\frac{y}{a}\right)\,\sin\left(n_z\,\pi\,\frac{z}{a}\right). \tag{12.97}$$

As is the case for a particle trapped in a one-dimensional potential well, the lowest energy level for a particle trapped in a three-dimensional well is not zero, but rather

$$E_{1,1,1} = 3\,E_1. \tag{12.98}$$

Here,

$$E_1 = \frac{\hbar^2}{2\,m\,a^2}. \tag{12.99}$$

is the *ground state* (*i.e.*, the lowest energy state) energy in the one-dimensional case. Now, it is clear, from (12.96), that distinct permutations of n_x, n_y, and n_z which do not alter the value of $n_x^{\,2}+n_y^{\,2}+n_z^{\,2}$ also do not alter the energy. In other words, in three dimensions it is possible for distinct wavefunctions to be associated with the same energy level. In this situation, the energy level is said to be *degenerate*. The ground state energy level, $3\,E_1$, is non-degenerate, since the only combination of (n_x, n_y, n_z) which gives this energy is (1, 1, 1). However, the next highest energy level, $6\,E_1$, is degenerate, since it is obtained when (n_x, n_y, n_y) take the values (2, 1, 1), or (1, 2, 1), or (1, 1, 2). In fact, it is not difficult to see that a non-degenerate energy level corresponds to a case where the three *mode numbers* (*i.e.*, n_x, n_y, and n_z) all have the same value, whereas a three-fold degenerate energy level corresponds to a case where only two of the mode numbers have the same value, and, finally, a six-fold degenerate energy level corresponds to a case where the mode numbers are all different.

12.12 Particle in a Finite Potential Well

Consider, now, a particle of mass m trapped in a one-dimensional square potential well of width a and finite depth $V > 0$. In fact, suppose that the potential takes the form

$$U(x) = \begin{cases} -V & |x| \leq a/2 \\ 0 & \text{otherwise} \end{cases}. \tag{12.100}$$

Here, we have adopted the standard convention that $U(x) \to 0$ as $|x| \to \infty$. This convention is useful because, just like in classical mechanics, a particle whose overall energy, E, is negative is bound in the well (*i.e.*, it cannot escape to infinity), whereas a particle whose overall energy is positive is unbound. Since we are interested in bound particles, we shall assume that $E < 0$. We shall also assume that $E + V > 0$, in order to allow the particle to have a positive kinetic energy inside the well.

Let us search for a stationary state

$$\psi(x,t) = \psi(x)\, e^{-i\,(E/\hbar)\,t}, \tag{12.101}$$

whose stationary wavefunction, $\psi(x)$, satisfies the time independent Schrödinger equation, (12.62). Now, it is easily appreciated that the solutions to (12.62) in the symmetric [*i.e.*, $U(-x) = U(x)$] potential (12.100) must be either totally symmetric [*i.e.*, $\psi(-x) = \psi(x)$] or totally antisymmetric [*i.e.*, $\psi(-x) = -\psi(x)$]. Moreover, the solutions must satisfy the boundary condition

$$\psi \to 0 \quad \text{as} \quad |x| \to \infty, \tag{12.102}$$

otherwise they would not correspond to bound states.

Let us, first of all, search for a totally symmetric solution. In the region to the left of the well (*i.e.*, $x < -a/2$), the solution of the time independent Schrödinger equation which satisfies the boundary condition $\psi \to 0$ as $x \to -\infty$ is

$$\psi(x) = A\, e^{\,q\,x}, \tag{12.103}$$

where

$$q = \sqrt{\frac{2\,m\,(-E)}{\hbar^2}}, \tag{12.104}$$

and A is a constant. By symmetry, the solution in the region to the right of the well (*i.e.*, $x > a/2$) is

$$\psi(x) = A\, e^{-q\,x}. \tag{12.105}$$

The solution inside the well (*i.e.*, $|x| \leq a/2$) which satisfies the symmetry constraint $\psi(-x) = \psi(x)$ is

$$\psi(x) = B\,\cos(k\,x), \tag{12.106}$$

where

$$k = \sqrt{\frac{2\,m\,(V+E)}{\hbar^2}}, \tag{12.107}$$

and B is a constant. The appropriate matching conditions at the edges of the well (*i.e.*, $x = \pm a/2$) are that $\psi(x)$ and $d\psi(x)/dx$ both be *continuous* [since a discontinuity in the wavefunction, or its first derivative, would generate a singular term in the time independent Schrödinger equation (*i.e.*, the term involving $d^2\psi/dx^2$) which could not be balanced]. The matching conditions yield

$$q = k\,\tan(k\,a/2). \tag{12.108}$$

Let $y = k\,a/2$. It follows that

$$E = E_0\,y^2 - V, \tag{12.109}$$

where

$$E_0 = \frac{2\,\hbar^2}{m\,a^2}. \tag{12.110}$$

Moreover, Equation (12.108) becomes

$$\frac{\sqrt{\lambda - y^2}}{y} = \tan y, \tag{12.111}$$

with

$$\lambda = \frac{V}{E_0}. \tag{12.112}$$

Here, y must lie in the range $0 < y < \sqrt{\lambda}$, in order to ensure that E lies in the range $-V < E < 0$.

Now, the solutions of Equation (12.111) correspond to the intersection of the curve $\sqrt{\lambda - y^2}/y$ with the curve $\tan y$. Figure 12.6 shows these two curves plotted for a particular value of λ. In this case, the curves intersect twice, indicating the existence of two totally symmetric bound states in the well. Moreover, it is clear, from the figure, that as λ increases (*i.e.*, as the well becomes deeper) there are more and more bound states. However, it is also apparent that there is always at least one totally symmetric bound state, no matter how small λ becomes (*i.e.*, no matter how shallow the well

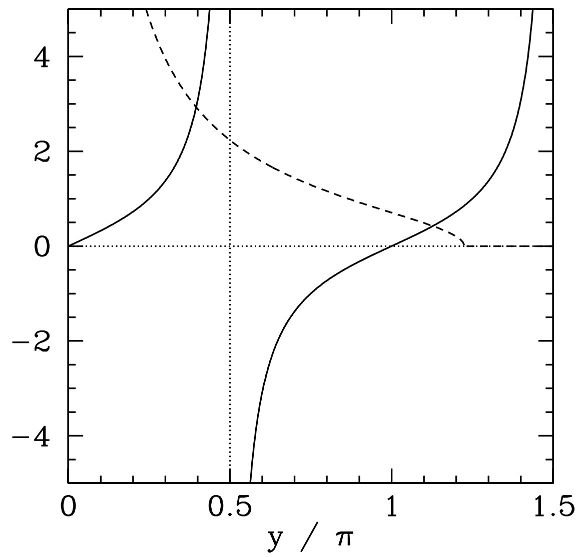


Figure 12.6: *The curves* $\tan y$ *(solid) and* $\sqrt{\lambda - y^2}/y$ *(dashed), calculated for* $\lambda = 1.5\,\pi^2$. *The latter curve takes the value* 0 *when* $y > \sqrt{\lambda}$.

becomes). In the limit $\lambda \gg 1$ (*i.e.*, the limit in which the well is very deep), the solutions to Equation (12.111) asymptote to the roots of $\tan y = \infty$. This gives $y = (2\,n - 1)\,\pi/2$, where n is a positive integer, or

$$k = (2\,n - 1)\,\frac{\pi}{a}. \tag{12.113}$$

These solutions are equivalent to the odd-n infinite-depth potential well solutions specified by Equation (12.68).

For the case of a totally antisymmetric bound state, similar analysis to the above yields (see Exercise 12.3)

$$-\frac{y}{\sqrt{\lambda - y^2}} = \tan y. \tag{12.114}$$

The solutions of this equation correspond to the intersection of the curve $\tan y$ with the curve $-y/\sqrt{\lambda - y^2}$. Figure 12.7 shows these two curves plotted for the same value of λ as that used in Figure 12.6. In this case, the curves intersect once, indicating the existence of a single totally antisymmetric bound state in the well. It is, again, clear, from the figure, that as λ increases (*i.e.*, as the well becomes deeper) there are more and more bound states. However, it is also apparent that when λ becomes sufficiently small

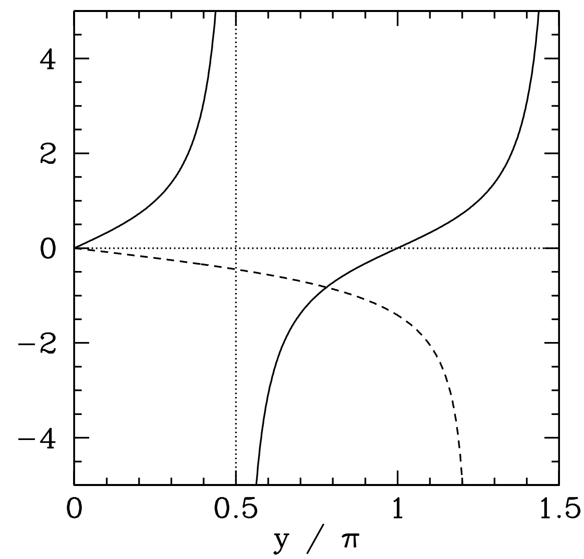


Figure 12.7: *The curves* $\tan y$ *(solid) and* $-y/\sqrt{\lambda - y^2}$ *(dashed), calculated for* $\lambda = 1.5\,\pi^2$.

[*i.e.*, $\lambda < (\pi/2)^2$] then there is no totally antisymmetric bound state. In other words, a very shallow potential well always possesses a totally symmetric bound state, but does not generally possess a totally antisymmetric bound state. In the limit $\lambda \gg 1$ (*i.e.*, the limit in which the well becomes very deep), the solutions to Equation (12.114) asymptote to the roots of $\tan y = 0$. This gives $y = n\,\pi$, where n is a positive integer, or

$$k = 2\,n\,\frac{\pi}{a}. \tag{12.115}$$

These solutions are equivalent to the even-n infinite-depth potential well solutions specified by Equation (12.68).

Probably the most surprising aspect of the bound states that we have just described is the possibility of finding the particle *outside* the well: *i.e.*, in the region $|x| > a/2$ where $U(x) > E$. This follows from Equation (12.105) and (12.106) because the ratio $A/B = \exp(q\,a/2)\,\cos(k\,a/2)$ is not necessarily zero. Such behavior is strictly forbidden in classical mechanics, according to which a particle of energy E is restricted to regions of space where $E > U(x)$. In fact, in the case of the ground state (*i.e.*, the lowest energy symmetric state) it is possible to demonstrate that the probability of a measurement finding the particle outside the well is (see Exercise 12.4)

$$P_{\rm out} \simeq 1 - 2\,\lambda \tag{12.116}$$

for a shallow well (*i.e.*, $\lambda \ll 1$), and

$$P_{\rm out} \simeq \frac{\pi^2}{4}\,\frac{1}{\lambda^{3/2}} \tag{12.117}$$

for a deep well (*i.e.*, $\lambda \gg 1$). It follows that the particle is very likely to be found outside a shallow well, and there is a small, but finite, probability of it being found outside a deep well. In fact, the probability of finding the particle outside the well only goes to zero in the case of an infinitely deep well (*i.e.*, $\lambda \rightarrow \infty$).

12.13 Square Potential Barrier

Consider a particle of mass m and energy $E > 0$ interacting with the simple potential barrier

$$U(x) = \left\{ \begin{array}{lll} V & & \mbox{for } 0 \leq x \leq a \\ 0 & & \mbox{otherwise} \end{array} \right., \tag{12.118}$$

where $V > 0$. In the regions to the left and to the right of the barrier, the stationary wavefunction, $\psi(x)$, satisfies

$$\frac{d^2\psi}{dx^2} = -k^2\,\psi, \tag{12.119}$$

where

$$k = \sqrt{\frac{2\,m\,E}{\hbar^2}}. \tag{12.120}$$

Let us adopt the following solution of the above equation to the left of the barrier (*i.e.*, $x < 0$):

$$\psi(x) = e^{\,i\,k\,x} + R\,e^{-i\,k\,x}. \tag{12.121}$$

This solution consists of a plane wave of unit amplitude traveling to the right [since the full wavefunction is multiplied by a factor $\exp(-i\,E\,t/\hbar)$], and a plane wave of complex amplitude R traveling to the left. We interpret the first plane wave as an *incoming particle*, and the second as a particle *reflected* by the potential barrier. Hence, $|R|^2$ is the probability of reflection (see Section 7.6).

Let us adopt the following solution to Equation (12.119) to the right of the barrier (*i.e.* $x > a$):

$$\psi(x) = T\,e^{i\,k\,x}. \tag{12.122}$$

This solution consists of a plane wave of complex amplitude T traveling to the right. We interpret this as a particle *transmitted* through the barrier. Hence, $|T|^2$ is the probability of transmission.

Let us consider the situation in which $E < V$. In this case, according to classical mechanics, the particle is unable to penetrate the barrier, so the coefficient of reflection is unity, and the coefficient of transmission zero. However, this is not necessarily the case in wave mechanics. In fact, inside the barrier (*i.e.*, $0 \leq x \leq a$), $\psi(x)$ satisfies

$$\frac{d^2\psi}{dx^2} = q^2\,\psi, \tag{12.123}$$

where

$$q = \sqrt{\frac{2\,m\,(V-E)}{\hbar^2}}. \tag{12.124}$$

The general solution to Equation (12.123) takes the form

$$\psi(x) = A\,e^{\,q\,x} + B\,e^{-q\,x}. \tag{12.125}$$

Now, continuity of ψ and $d\psi/dx$ at the left edge of the barrier (*i.e.*, $x = 0$) yields

$$1 + R = A + B, \tag{12.126}$$

$$i\,k\,(1-R) = q\,(A-B). \tag{12.127}$$

Likewise, continuity of ψ and $d\psi/dx$ at the right edge of the barrier (*i.e.*, $x = a$) gives

$$A\,e^{\,q\,a} + B\,e^{-q\,a} = T\,e^{\,i\,k\,a}, \tag{12.128}$$

$$q\left(A\,e^{\,q\,a} - B\,e^{-q\,a}\right) = i\,k\,T\,e^{\,i\,k\,a}. \tag{12.129}$$

After considerable algebra (see Exercise 12.5), the above four equations yield

$$|T|^2 = 1 - |R|^2 = \frac{4\,k^2\,q^2}{4\,k^2\,q^2 + (k^2+q^2)^2\,\sinh^2(q\,a)}. \tag{12.130}$$

Here, $\sinh x \equiv (1/2)\,(e^{\,x} - e^{-x})$. The fact that $|R|^{\,2} + |T|^{\,2} = 1$ ensures that the probabilities of reflection and transmission sum to unity, as must be the case, since reflection and transmission are the only possible outcomes for a particle incident on the barrier. Note that, according to Equation (12.130), the probability of transmission is not necessarily zero. This means that, in wave mechanics, there is a finite probability for a particle incident on a potential barrier, of finite width, to penetrate through the barrier, and reach the other side, even when the barrier is sufficiently high to completely reflect the particle according to the laws of classical mechanics. This strange phenomenon is known as *tunneling*. For the case of a very high barrier, such that $V \gg E$, the tunneling probability reduces to

$$|T|^{\,2} \simeq \frac{4\,E}{V}\,e^{-2\,a/\lambda}, \tag{12.131}$$

where $\lambda = \sqrt{\hbar^2/2\,m\,V}$ is the de Broglie wavelength inside the barrier. Here, it is assumed that $a \gg \lambda$. Note that, even in the limit in which the barrier is very high, there is an exponentially small, but nevertheless *non-zero*, tunneling probability. Tunneling plays an important role in the physics of α-decay and electron field emission.

12.14 Exercises

1. Use the standard power law expansions,

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, \\ \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots, \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots, \end{aligned}$$

which are valid for complex x, to prove de Moivre's theorem,

$$e^{\,i\,\theta} = \cos\theta + i\,\sin\theta,$$

where θ is real.

2. Equations (8.27) and (8.28) can be combined with de Moivre's theorem to give

$$\delta(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\,i\,k\,x}\,dx,$$

where $\delta(k)$ is a Dirac delta function. Use this result to prove Fourier's theorem: *i.e.*, if

$$f(x) = \int_{-\infty}^{\infty} \bar{f}(k)\, e^{\,i\,k\,x}\, dk,$$

then

$$\bar{f}(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\, e^{-i\,k\,x}\, dx.$$

3. Derive Equation (12.114).

4. Consider a particle trapped in the finite potential well whose potential is given by Equation (12.100). Demonstrate that for a totally symmetric state the ratio of the probability of finding the particle outside to the probability of finding the particle inside the well is

$$\frac{P_{\rm out}}{P_{\rm in}} = \frac{\cos^3 y}{\sin y\,(y + \sin y\, \cos y)},$$

where $\sqrt{\lambda - y^2} = y\,\tan y$, and $\lambda = V/E_0$. Hence, demonstrate that for a shallow well (*i.e.*, $\lambda \ll 1$) $P_{\rm out} \simeq 1 - 2\,\lambda$, whereas for a deep well (*i.e.*, $\lambda \gg 1$) $P_{\rm out} \simeq (\pi^2/4)/\lambda^{3/2}$.

5. Derive expression (12.130) from Equations (12.126)–(12.129).

6. Show that the coefficient of transmission of a particle of mass m and energy E, incident on a square potential barrier of height $V < E$, and width a, is

$$|T|^2 = \frac{4\,k^2\,q^2}{4\,k^2\,q^2 + (k^2 - q^2)^2\,\sin^2(q\,a)},$$

where $k = \sqrt{2\,m\,E/\hbar^2}$ and $q = \sqrt{2\,m\,(E - V)/\hbar^2}$. Demonstrate that the coefficient of transmission is unity (*i.e.*, there is no reflection from the barrier) when $q\,a = n\,\pi$, where n is positive integer.

7. A He-Ne laser emits radiation of wavelength $\lambda = 633$ nm. How many photons are emitted per second by a laser with a power of 1 mW? What force does such a laser exert on a body which completely absorbs its radiation?

8. The ionization energy of a hydrogen atom in its ground state is $E_{\rm ion} = 13.6$ eV. Calculate the frequency (in Hertz), wavelength, and wavenumber of the electromagnetic radiation which will just ionize the atom.

9. The maximum energy of photoelectrons from aluminium is 2.3 eV for radiation of wavelength 200 nm, and 0.90 eV for radiation of wavelength 258 nm. Use this data to calculate Planck's constant (divided by 2π) and the work function of aluminium.

10. Show that the de Broglie wavelength of an electron accelerated across a potential difference V is given by

$$\lambda = 1.29 \times 10^{-9}\, V^{-1/2}\, \text{m},$$

where V is measured in volts.

11. If the atoms in a regular crystal are separated by 3×10^{-10} m demonstrate that an accelerating voltage of about 3 kV would be required to produce an electron diffraction pattern from the crystal.

12. A particle of mass m has a wavefunction

$$\psi(x,t) = A\, \exp\left[-a\,(m\,x^2/\hbar + i\,t)\right],$$

where A and a are positive real constants. For what potential $U(x)$ does $\psi(x,t)$ satisfy Schrödinger's equation?

13. Show that the wavefunction of a particle of mass m trapped in a one-dimensional square potential well of of width a, and infinite depth, returns to its original form after a quantum revival time $T = 4\,m\,a^2/\pi\,\hbar$.

14. Show that the normalization constant for the stationary wavefunction

$$\psi(x,y,z) = A\, \sin\left(n_x\,\pi\,\frac{x}{a}\right)\sin\left(n_y\,\pi\,\frac{y}{b}\right)\sin\left(n_z\,\pi\,\frac{z}{c}\right)$$

describing an electron trapped in a three-dimensional rectangular potential well of dimensions a, b, c, and infinite depth, is $A = (8/abc)^{1/2}$. Here, n_x, n_y, and n_z are positive integers.

15. An electron of momentum p passes through a slit of width Δx. Its diffraction as a wave can be regarded in terms of a change of its momentum Δp in a direction parallel to the plane of the slit (the total momentum remaining constant). Show that the approximate position of the first maximum of the diffraction pattern is in accordance with Heisenberg's uncertainty principle.

16. The probability of a particle of mass m penetrating a distance x into a classically forbidden region is proportional to $e^{-2\,\alpha\,x}$, where

$$\alpha^2 = 2\,m\,(V - E)/\hbar^2.$$

If $x = 2 \times 10^{-10}$ m and $V - E = 1$ eV show that $e^{-2\,\alpha\,x}$ is equal to 0.1 for an electron, and 10^{-43} for a proton.

A Useful Information

A.1 Physical Constants

Speed of light	c	2.998×10^{8}	$\mathrm{m\,s^{-1}}$
Reduced Planck constant	$\hbar$	1.055×10^{-34}	$\mathrm{J\,s}$
Electric constant	ϵ_0	8.854×10^{-12}	$\mathrm{F\,m^{-1}}$
Magnetic constant	μ_0	1.257×10^{-6}	$\mathrm{N\,A^{-2}}$
Elementary charge	e	1.602×10^{-19}	C
Electron mass	m_e	9.109×10^{-31}	kg
Proton mass	m_p	1.673×10^{-27}	kg

A.2 Trigonometric Identities

$$
\begin{aligned}
\sin(-\alpha) &= -\sin\alpha \\
\cos(-\alpha) &= +\cos\alpha \\
\tan(-\alpha) &= -\tan\alpha \\
\sin^2\alpha + \cos^2\alpha &= 1 \\
\sin(\alpha \pm \beta) &= \sin\alpha\,\cos\beta \pm \cos\alpha\,\sin\beta \\
\cos(\alpha \pm \beta) &= \cos\alpha\,\cos\beta \mp \sin\alpha\,\sin\beta \\
\sin\alpha + \sin\beta &= 2\,\sin\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right) \\
\sin\alpha - \sin\beta &= 2\,\cos\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right) \\
\cos\alpha + \cos\beta &= 2\,\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right) \\
\cos\alpha - \cos\beta &= -2\,\sin\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right) \\
\sin\alpha\,\sin\beta &= \frac{1}{2}\left[\cos(\alpha-\beta) - \cos(\alpha+\beta)\right] \\
\cos\alpha\,\cos\beta &= \frac{1}{2}\left[\cos(\alpha-\beta) + \cos(\alpha+\beta)\right] \\
\sin\alpha\,\cos\beta &= \frac{1}{2}\left[\sin(\alpha-\beta) + \sin(\alpha+\beta)\right]
\end{aligned}
$$

$$
\begin{aligned}
\sin(2\,\alpha) &= 2\,\sin\alpha\,\cos\alpha \\
\cos(2\,\alpha) &= \cos^2\alpha - \sin^2\alpha = 2\,\cos^2\alpha - 1 = 1 - 2\,\sin^2\alpha \\
\sin(3\,\alpha) &= -4\,\sin^3\alpha + 3\,\sin\alpha \\
\cos(3\,\alpha) &= 4\,\cos^3\alpha - 3\,\cos\alpha \\
\sin^2\alpha &= [1 - \cos(2\,\alpha)]/2 \\
\cos^2\alpha &= [1 + \cos(2\,\alpha)]/2
\end{aligned}
$$

A.3 Calculus

$$
\begin{aligned}
\frac{d}{dx}\,x^n &= n\,x^{n-1} \\
\frac{d}{dx}\,\sin x &= \cos x \\
\frac{d}{dx}\,\cos x &= -\sin x \\
\frac{d}{dx}\,e^x &= e^x \\
\frac{d}{dx}\,\ln x &= \frac{1}{x} \\
\frac{d}{dx}\,(f\,g) &= \frac{df}{dx}\,g + \frac{dg}{dx}\,f \\
\frac{d}{dx}\left(\frac{f}{g}\right) &= \frac{df}{dx}\,\frac{1}{g} - \frac{dg}{dx}\,\frac{f}{g^2} \\
\int f\,\frac{dg}{dx}\,dx &= f\,g - \int \frac{df}{dx}\,g\,dx
\end{aligned}
$$

A.4 Power Series

$$
\begin{aligned}
(1+x)^n &= 1 + n\,x + \frac{n\,(n-1)}{2!}\,x^2 + \cdots \\
e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \\
\ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots
\end{aligned}
$$

$$
\begin{aligned}
\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \\
\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots \\
\tan x &= x + \frac{x^3}{3} + \frac{2}{15}\,x^5 + \cdots \\
f(a+x) &= f(a) + f'(a)\,x + f''(a)\,\frac{x^2}{2!} + f'''(a)\,\frac{x^3}{3!} + \cdots
\end{aligned}
$$